LECTURES ON THE SCIENTIFIC BASIS OF MEDICINE

1958–59

British Postgraduate Medical Federation
University of London

LECTURES ON THE SCIENTIFIC BASIS OF MEDICINE

Volume VIII

1958–59

UNIVERSITY OF LONDON
THE ATHLONE PRESS
1960

PREFACE

NEW methods for the investigation of biological problems are being constantly evolved and many of these can be applied to the elucidation of the problems met with in medicine or modified to meet the peculiar conditions of clinical investigations. It is, therefore, essential that clinicians should be kept informed of the progress being made in the biological sciences and especially the sciences on which the practice of medicine is based. For this reason and because no clinician or research worker has the time to keep in touch with the progress that is taking place in the branches of science in which he is not himself actively concerned, the British Postgraduate Medical Federation arranges a series of lectures each winter on subjects that illustrate the use of new methods and of the progress taking place. The lecturers are themselves actively engaged in research and the lectures are designed for graduates in medicine and dentistry looking forward to careers as clinical specialists as well as for those seeking careers in research in the medical sciences. This volume, the eighth in the series, contains twenty-four out of the thirty lectures delivered during the winter of 1958–59. The lateness of the volume in appearing gives cause for regret but, in company with many other scientific publications of a serial nature, it has suffered from the effects of the stoppage in the printing industry that occurred in the summer of 1959.

The importance of the cell and its constituent structures for an understanding of health and disease is exemplified in this volume by the lectures by Mercer on electron microscopy, by Abercrombie on the control of growth and the cell surface, and by Astbury on fine structural studies of the collagen-apatite partnership; some genetic aspects of cytology are discussed by Hayes on bacterial genetics and gene structure and Armstrong on intersexuality, while Scowen describes his studies on hyperoxaluria, establishing this condition as an inborn error of metabolism. Some outstanding problems in bacteriology and

Published by
THE ATHLONE PRESS
UNIVERSITY OF LONDON
at 2 Gower Street, London WC1
Distributed by Constable & Co Ltd
12 Orange Street, London WC2

Canada
University of Toronto Press

U.S.A.
Oxford University Press Inc
New York

Printed in Great Britain by
WESTERN PRINTING SERVICES LTD
BRISTOL

CONTENTS

immunology are dealt with by Stuart-Harris in his lecture on the adeno-viruses and respiratory disease, by Williams on epidemic staphylococcal infections in hospitals, Evans on vaccination against poliomyelitis and by Humphreys in his paper on gamma globulins.

Maizels contributes an historical analysis entitled 'The Role of Biochemistry in Medicine', and other lectures on biochemical subjects are those by Wrong on sodium excretion and the control of extracellular fluid volume, by Walshe on biochemical studies in hepatic coma and by Jenkins on the biochemical background to the action of fluoride in dental caries. The cardiovascular system is represented by the lectures by Harrison on pulmonary hypertension, by Byrom on the significance of hypertensive encephalopathy and by Greenfield on the regulation of the blood vessels of the limbs. The applications of new methods to old problems in haematology are seen in the lectures by Prankerd on the viability and survival of red blood cells and by Cappell on some aspects of iron metabolism. Lynne Reid's lecture on chronic bronchitis and hypersecretion of mucus is based on histological studies, Mole's lecture on radiation as a toxic agent is a contribution to a subject of world-wide interest at the present time, and Cope's on the measurement of adrenal activity in man is a useful summary of a subject of fundamental importance. Miles's lecture on mediators of inflammation illustrates the essential importance of a critical assessment of the contributions to a problem, and Salisbury's on the influence of botany on health emphasises that medicine is but part of the wide science of biology.

A volume based on the lectures delivered during the winter of 1959–60 is now being prepared and, among other subjects, will again illustrate the intensive studies that are being made on cells and the micro-structures they contain, their chemical constitution, and their functions in the maintenance of health; in addition there will be further contributions on the effects of radiations on biological processes.

FRANCIS R. FRASER
Director, British Postgraduate Medical Federation

1 July 1960

I

The Influence of Botany on Health and Disease

E. J. SALISBURY

I FIND myself today in the vanguard of a series of lectures on a diversity of most interesting topics by very distinguished lecturers. May I say that as an ecologist I find it particularly gratifying to see the catholicity of interests that those attending these lectures will be privileged to enjoy. Whilst most dealt with important aspects of the physiology and biochemistry of the organism, whether in health or disease, others are concerned with more general topics such as cell surface and the control of growth, the structure of the cell as revealed by the electron microscope or the genetics of bacteria.

As most of my audience will be aware, the effective practice of the healing art involves the assessment of many imponderables, as well as the application of all that meticulous scientific investigation can suggest. It cannot be too strongly emphasized that the increasing specialization which is an inevitable precursor of the advancement of knowledge, can only yield its full benefits if apprehended in correct evaluation against the background of biological knowledge as a whole. Wisdom is shown in the blending of the general and the particular in these lectures, realizing the validity of the view expressed by Sir Harold Himsworth, in his Linacre Lecture, that 'If medicine is to maintain its unity then it must include among its members, not only the whole range of specialists in its component subjects but men who, within their own persons, link relevant disciplines.'

It is because I believe you would endorse such sentiments that

NOTE

The lectures printed in this volume were delivered on the following dates:

I. 16 October 1958	XIII. 27 January 1959
II. 27 November 1958	XIV. 17 February 1959
III. 25 November 1958	XV. 29 January 1959
IV. 2 December 1958	XVI. 24 February 1959
V. 19 February 1959	XVII. 15 January 1959
VI. 20 November 1958	XVIII. 12 March 1959
VII. 18 November 1958	XIX. 5 March 1959
VIII. 23 October 1958	XX. 4 December 1958
IX. 21 October 1958	XXI. 3 February 1959
X. 30 October 1958	XXII. 10 February 1959
XI 22 January 1959	XXIII. 14 January 1959
XII. 3 March 1959	XXIV. 11 March 1959

of the active constituents are made by others we are perhaps apt to forget that the need for meticulous identification still remains, and has become not less but more important, as the intensive study of superficially similar plants reveals the existence of more active and less active strains and the biochemist discovers that the physiologically significant constituents are not one but several, present in varying proportions in the different populations that go to make that aggregate of convenience that we term a species.

The desirability of the medical practitioner himself knowing a number of plants nevertheless remains, since poisonous wild or introduced species still continue to take their toll of human lives. I recall that during the last world war I was sent some tuber-like roots for identification because similar ones had been added by Italian prisoners to their stewpot when engaged in cleaning out a ditch. They were the roots of the Water Dropwort (*Oenanthe crocata*) and not unnaturally the consequences were fatal. This plant was known to our forefathers as Dead-tongue because of the paralysing effect on the organs of speech that it can produce. Fortunately, however, it is more often responsible for the poisoning of cattle than of human beings The Deadly Nightshade (*Atropa belladonna*) is a plant everyone should know by sight since it continues, almost annually, to account for illness, sometimes fatal, amongst children attracted by its glistening-black cherry-like berries. So too children playing with the pods of the Laburnum Tree (*Laburnum anagyroides*) have not infrequently eaten them with fatal results. The fruits of other conspicuous wild plants such as the Black Bryony (*Tamus communis*), the Black Nightshade (*Solanum nigrum*), the Woody Nightshade (*Solanum dulcamara*), the Wild Arum (*Arum maculatum*) and those of the Yew, all continue to contribute their quota of malaise if not worse. The pink aril of the Yew fruit which is sweet and mucilaginous is itself innocuous, but as children are very liable to swallow the stones which contain about 0 08 per cent of the very poisonous alkaloid taxin, deaths from this cause occur from time to time. All these might usefully find a place in school gardens for the edification of the pupils. Probably the largest number of actual deaths from the eating of poisonous

I venture to address you on a subject that is somewhat peripheral to your main interests.

The spectacular progress of medical science in the fields of antitoxins, chemotherapy and other directions has had a profound influence upon the control of death. So much lower is juvenile mortality than a century ago that anyone who has given thought to the matter must realize that the future happiness of the human race and the avoidance of major catastrophes, including global war, will ultimately depend upon control of births matching our control of infant mortality. But despite the diminished severity of selection, which in the past usually only permitted the more robust to survive, the improved hygiene and generally better standards of life and livelihood have added some ten years to the expectation of the human span. There are still serious diseases, such as rheumatic arthritis and malignant tumours, which challenge our efforts for alleviation and as yet elude control, but, if we are to live longer, that will be no incubus placed upon the community provided the maladies of old age are kept at bay and decreasing vigour, both physical and mental, is proportionately postponed. As I pointed out some years ago when opening a discussion on the span of life at the Royal Society of Medicine, there is much indirect evidence to suggest that senescence is an accompaniment of the increasing accumulation of the by-products of metabolic activity and if all that has just been stated be true, it is apparent that the emphasis of the future will be on prevention and amelioration rather than on cure. It is in this context that I would particularly wish to direct your attention to the contribution that botany can make to the problems of health and disease and the betterment of human welfare.

It is unnecessary for me to remind you that the study of plants originated from the human compulsion to distinguish between the nutritious and the harmful and to recognize the simples by means of which primitive man alleviated his maladies. The subject of botany as an integral part of the medical curriculum was an essential element in the practical training when the doctor himself was under the necessity of identifying and preparing the herbs which he employed. Now that the preparations

develops the appearance of suffering from an attack of measles, presents us with problems that are biochemical rather than botanical unless there should prove to be interspecific differences to be elucidated. The incapacitating and all too frequent allergies, known as Hay-fever, have, however, been the subject of fruitful botanical investigation from the aspect of the species of plant whose pollen is responsible and the seasonal incidence of its wind-borne dispersal.

The so-called pollen spectrum differs from one country to another and to a lesser degree from district to district. In the United States, for example, three main periods of occurrence can be recognized. The spring phase, in which pollen of such trees as Willows, Maples, Birches and Oaks is significant; the summer phase, dominated by the pollen of Grasses and Sedges; and the autumn phase in which the pollen of Ragweeds (*Ambrosia* spp.) and Golden-Rods (*Solidago* spp.) are the chief irritants. The studies made by Dr. Hyde in this country upon air samples, from various localities throughout the year, indicate the preponderance of Grass pollen during June and July, a significant occurrence of Tree pollen during March, April and May, but only a relatively small amount of pollen in the autumn, consisting mainly of Mugwort (*Artemesia vulgaris*) and Chenopodiaceae. It is nevertheless important to remember that, except for highly sensitized victims, the kind of pollen may be far more important than the amount, and a particular individual may be specially allergic to the antigens of a species whose pollen is less frequently present in the atmosphere. The pollen of Coniferous trees though locally produced in great quantities in spring is stated to be not normally a significant cause of allergic response, although the pollen of Cedars appears to be an exception, whilst Elm pollen though it may be small in amount may be drastic in its effect upon some individuals.

To those who might doubt the continuing practical value to medicine of intensive botanical studies in the taxonomic domain, the example of *Penicillin* is I think instructive. As you are doubtless aware there are a very large number of species of the blue moulds that are placed in this genus, of which most, if not all, produce antibiotics. But only two appear to yield an antibiotic

plants are attributable to consumption of the 'Death Cap' toadstool (*Amanita phalloides*) in place of mushrooms, despite the easy differentiation of the two by the colour of the gills which are white or cream-coloured in the Death Cap instead of pink in the young mushroom or purplish-brown in the mature one. If only people were more epicurean and ate mushrooms when they had expanded and developed their full flavour instead of eating them as mere buttons when they are relatively flavourless, because they look attractive, there would be far fewer such mistakes. Owing to the high toxicity and the lapse of from six to fifteen hours before marked symptoms develop after eating the Death Cap, it is then often too late to apply remedial measures.

There are those who advocate the destruction of all poisonous plants, but the aim should be education rather than eradication. Attempts at eradication are rarely wholly successful, and by thus rendering the species concerned rare and unfamiliar such action is calculated to defeat the greater safeguard. More comprehensible, though now fortunately infrequent, is the unpleasant contact poisoning by the American Poison Ivy (*Rhus toxicodendron*) occasionally planted in mistake for Virginia Creeper. Some years ago a house at Watford achieved a poltergeist reputation which was only dispelled when a knowledgeable local practitioner recognized the offending climber and exorcised the evil presence, after which the maids were able to wave out of the windows to their young men without untoward consequences.

The mistaken substitution of Aconite roots for Horseradish and of Fool's Parsley for the real Parsley is now rarely perpetrated, but this brief reference to a few of the plants most frequently causing illness or death perhaps justifies the assertion that some knowledge of plants is still a desirable medical qualification, quite apart from the fact that if the practitioner is to inspire confidence in his patients he must be well educated and a modicum of botanical lore is, I suggest, an essential part of a liberal education. The familiar allergies are a constant reminder of the old adage that 'one man's meat is another man's poison'. For example, most of us, luckily, can eat strawberries with impunity, but the fact that an occasional individual, even after eating only one or two of these fruits, quickly

by the Almighty, it was natural piety to expect that its structure and habit would have some significance and probably be indicative of the purpose of its creation. It is no exaggeration to assert that the search for such indications, by the pious seeker after truth, was the means of adding greatly to the knowledge of herbs. Indeed no better example could be given of the fact that *sincerity of purpose in its pursuit is more important for the advancement of knowledge than the truth of the hypothesis that stimulates investigation.* The transmutation of elements that the alchemists believed in, and for which they were long derided, has proved to be true in a way they could not dream of, but those who pursued their purpose with the wholehearted desire for truth laid the foundation of modern chemistry and so too the sincere herbalists, actuated by their pious beliefs, laid the foundations of a botanical superstructure that developed into the pharmacognoscy of today. No one can, I think, doubt that one important contribution of botany to the promotion of health has been the cumulative diagnosis of plant characteristics that has made the assessment of their resemblances and differences increasingly assured, so that the subjective aggregates that we term families, genera or species come to correspond more and more closely to their genetical relationships. The cytologists have added further features in the numbers and form of the chromosomes that have sometimes modified the older assessments but, in the main, they have confirmed them, whilst the soundness of taxonomic judgements in general has been attested by the results of serological experiments. Resemblances of structure, whether external or internal, macroscopic or microscopic, have proved a valuable guide to affinity and thus often the visible has been a guide to the more important invisible physiological characteristics, including the metabolic by-products such as those employed as drugs, which often can only be ascertained by prolonged experimentation. It is therefore no idle claim that the work of the taxonomic botanist has been of great importance to medical science.

It has been asserted that the number of therapeutic drugs that are really essential is very limited. No doubt opinions would vary greatly as to what should be included in any such list but there

that is at once highly toxic to bacteria and almost innocuous to man. On the other hand the commonest and ubiquitous species *Penicillium glaucum* produces an antibiotic that is worse than useless therapeutically, since it is harmful to man as well as to bacteria. It is clear that if botanists had not learnt to distinguish these various species of *Penicillium* the task which Sir Howard Florey so successfully accomplished would have been rendered far more difficult, if not insuperable. It is I think salutary to speculate on the answer that would probably have been accorded fifty years ago to the question whether the study of the various blue moulds, on food and other organic substrata, was a proper occupation to engross the attention of an intelligent person. I rather suspect that the most tolerant answer would have been either impolite or that it was a harmless pastime for one who could find nothing better on which to squander his leisure.

I suggest, however, that it may help us to place botanical identification in its modern context in relation to medical science if we consider briefly a few examples where the therapeutic importance is associated with sub-specific distinctions. The genetic studies of modern times have shown that within the species there are strains that exhibit a wide range of variation in their capacity to produce the active constituents and both the inherited nature of the plant as well as its nurture combine to determine the measure of its physiological activity.

Apart from these distinctions within the species, however, it is worth noting how few plants of medicinal importance have been brought to light that had not already been employed by our forefathers, a fact that bears witness to the tribulations of primitive man and the long experience that produced the empirical knowledge as to what plants were active. In this connection I should like to refer to the 'doctrine of signatures' which, though it attained in the hands of some of its disciples an absurd extravagance, played, I venture to think, a very important part in the progress of the lore of simples. Stated briefly the underlying conception—doubtless long antedating any recognition of a formal doctrine—was that since what was in the world should be regarded as having been created for the benefit of man

by the Almighty, it was natural piety to expect that its structure and habit would have some significance and probably be indicative of the purpose of its creation. It is no exaggeration to assert that the search for such indications, by the pious seeker after truth, was the means of adding greatly to the knowledge of herbs. Indeed no better example could be given of the fact that *sincerity of purpose in its pursuit is more important for the advancement of knowledge than the truth of the hypothesis that stimulates investigation.* The transmutation of elements that the alchemists believed in, and for which they were long derided, has proved to be true in a way they could not dream of, but those who pursued their purpose with the wholehearted desire for truth laid the foundation of modern chemistry and so too the sincere herbalists, actuated by their pious beliefs, laid the foundations of a botanical superstructure that developed into the pharmacognoscy of today. No one can, I think, doubt that one important contribution of botany to the promotion of health has been the cumulative diagnosis of plant characteristics that has made the assessment of their resemblances and differences increasingly assured, so that the subjective aggregates that we term families, genera or species come to correspond more and more closely to their genetical relationships. The cytologists have added further features in the numbers and form of the chromosomes that have sometimes modified the older assessments but, in the main, they have confirmed them, whilst the soundness of taxonomic judgements in general has been attested by the results of serological experiments. Resemblances of structure, whether external or internal, macroscopic or microscopic, have proved a valuable guide to affinity and thus often the visible has been a guide to the more important invisible physiological characteristics, including the metabolic by-products such as those employed as drugs, which often can only be ascertained by prolonged experimentation. It is therefore no idle claim that the work of the taxonomic botanist has been of great importance to medical science.

It has been asserted that the number of therapeutic drugs that are really essential is very limited. No doubt opinions would vary greatly as to what should be included in any such list but there

would perhaps be general agreement as to the indispensability of some, amongst which the plant products digitalin and morphine, and probably atropine also, would almost certainly be numbered. Seeds of the Opium Poppy (*Papaver somniferum*), the Deadly Nightshade (*Atropa belladonna*) and the Henbane (*Hyoscyamus niger*) have been found in deposits of the Roman period in Britain and were perhaps then all in use and cultivation. The home of the Opium Poppy is unknown, but it was employed by Egyptian physicians 1600 years before the Christian era and was probably known in the East earlier still. It was not however till 1816 that precision was given to the use of opium products through the isolation of the active principle by a Hanoverian apothecary.

When one considers the contribution that the study of plants has made to medical knowledge it must not be forgotten that most botanists until recent times were trained for the medical profession, so that botany owes at least as much to medicine as medicine to botany. The use of the Foxglove is an example. It was mentioned by the herbalists but was not held in great repute by all as the following passage from Johnson's Gerard of 1536 attests. Johnson wrote: 'The Fox-gloves in that they are bitter, are hot and dry, with a certaine kind of clensing qualitie joyned therewith; yet are they of no use, neither have they any place amongst medicines, according to the ancients.' It was, however, the distinguished physician and botanist William Withering who as an outcome of careful experimentation produced convincing evidence of the value of *Digitalis* for dropsy. This classical account published in 1776 is not only much sought after by bibliophiles but is still worth reading for its own sake. Withering's *Botanical Arrangement of the Vegetables in Great Britain*, a flora in two volumes, which also appeared in 1776, went through no less than seven editions, of which the last was issued in 1830, so that one may infer that his services to botany were on a par with those he rendered to medicine. The early herbalists, who did so much for plant identification, had little concept of what we understand by experiment, and the progress of empirical deduction was greatly hampered by the practice of incorporating so many ingredients in the same specific, presumably in the

hope that if one ingredient failed another would fulfil its role. The fifteenth-century Leechbook, which was edited by W. R. Dawson and published by the Royal Society of Literature, furnishes a good illustration of this. It describes a cough cure which contained Avens, Betony, Borage, Calamint, Chamomile, Cinnamon, Cloves and Comfrey, Daisy, Elecampane, Eyebright, Galingale, Ginger, Harts-tongue, Herb Robert, Hyssop, Laurel, Lavender, Liquorice, Mace, Marjoram, Milfoil, Mint, Mouse-Ear, Mugwort, Nutmeg, Radish, Rosemary, Sage, Savory, Scabious, Thyme, Tormentil, Valerian, Violet, Waybread and Wormwood. Thirty-seven constituents! Of most of these you were told to use a small handful except for the liquorice, probably the efficacious part, of which a pound or more was included. With such recipes it is small wonder that empiricism made little headway till simpler formulae became the vogue. Today we enjoy the benefits of the trials, which must often have been grievous, and the errors that were doubtless multitudinous, of past centuries. This accumulation of empirical lore has often been the starting-point from which exact experimentation has emanated. If modern science has converted the crude and often chancy medicaments of the past into present instruments of considerable precision, we should not belittle the contribution of acute observers who perceived the significant amidst the fumblings and superstitions of the pre-experimental era. These minds separated the true ore of facts from the dross of superstition and groundless speculation. Kipling indeed was guilty of little exaggeration when he wrote

Everything green that grew out of the mould
Was a wonderful herb to our fathers of old

Owing to the impossibility of obtaining the usual supplies of Henbane from the Continent during the last world war, other sources of hyoscin were exploited and it was found that the small Solanaceous tree *Duboisia hopwoodii*, which is chewed and smoked by the Australian natives as a stimulant and narcotic, contained an appreciable content of hyoscin and also the less desirable hyosciamin. But whereas the trees in Queensland gave a satisfactory yield further south the plants were found to be of

little value because they contained a high proportion of hyosciamin and little hyoscin. Since then this distinction has been shown to be a genetic one. The two strains are similar in appearance but their physiological distinction, which is doubtless responsible for the geographical segregation, is accompanied by this biochemical difference that entirely changes their medicinal importance.

Incidentally we may note that the search for substitutional drug sources has usually depended upon the exploitation of other plants of the same affinity, and since these are not infrequently wholly different in habit from their relatives it is manifest that their recognition as possible sources is dependent upon botanical research. We see this exemplified in relation to the drug reserpine obtained from the root of the Indian plant *Rauwolfia serpentina.* The story of the isolation of this active principle from a plant long used in Indian medicine is a striking illustration of the frequent dependence of advance upon the development of new techniques. When, by the employment of electrophoresis, the active constituents were finally separated from the Oleoresin residue, which the Sidiquis had not the technique to fractionate since this was only discovered some years later, a crude medicament was converted into a drug of precision and the demand became so great that an embargo had to be placed upon the export of the roots of the plant from India because of the danger of its extinction. This led to a search amongst other plants that botanists had placed in the same family of the Apocyanaceae and a plant from New Zealand very *unlike Rauwolfia was found to be a useful source of reserpine.*

I think one could justly claim that it was the manifestly deleterious effects of atmospheric pollution upon plant life which first brought into prominence the probable indirect effects of this accompaniment of the industrial revolution on human health. It is certainly a fact that the first really significant quantitative data as to the materials added to the atmosphere were obtained nearly seventy years ago by a botanist, Professor F. W. Oliver, at the instance of the Royal Horticultural Society. For our present context, however, our interest must be directed not so much to the soot and grosser constituents

of smoke as to the physiologically active gases. It is now over half a century since Neljubov demonstrated that as small a quantity as one part per million of the gas ethylene could bring about distortion in the growth of the plumule of the pea and thereby established, beyond equivocation, that the continued operation of physiologically active chemical substances, in amounts of extreme diminitude, could produce effects of profound significance.

I need scarcely remind this audience of the prejudice that had to be overcome at that period in the promulgation of such a conception, amongst physiologists whose ranks were then prevailingly medical, since the bitter controversy between the allopaths and the homoeopaths was still fresh in men's minds. Hanneman's doctrine of dynamization which he promulgated in the early years of the nineteenth century, linked as it was in the thoughts of orthodox practitioners with the more extravagant claims of homoeopathy, such as that like always cured like, was understandably looked upon askance by the allopaths and thus the notion that very minute quantities of any chemical substance could be efficacious in promoting human well-being was regarded as a concession to medical heretics.

Subsequent experiments have shown that some plants exhibit characteristic growth responses to as minute a trace as one part in ten million. Since ordinary illuminating gas often contains about 4 per cent of ethylene, this implies that a gas escape of 1 part in 400,000 of air could produce a physiological result. We know too that the minute quantities of ethylene produced during the ripening of apples, unless removed by ventilation, has a pronounced accelerating influence upon the maturation of immature fruits adjacent, in the same atmosphere. It is unnecessary to remind you that traces of ethylene are known to affect the human body and, *inter alia*, to promote an anaemic tendency. What I am, however, anxious to stress here is that these and similar discoveries in the botanical domain respecting minute quantities of ethylene and sulphur dioxide, prepared men's minds for the concept that major effects could be the outcome of infinitesimal causes, a fact later to be spectacularly demonstrated for human nutrition, in the domain of organic

chemistry, by Gowland Hopkins' discovery of the vitamins and, in the domain of inorganic chemistry, with respect to the trace elements or micronutrients that have been shown as essential for the nutrition of plants and animals alike.

Empirically, the value of plants containing vitamin C has probably been appreciated from ancient times. One of the plants so used in Britain was the Bistort (*Polygonum bistorta*) whose leaves were some of the earliest that could be induced to sprout after 'the winter of discontent' that our forefathers endured upon a diet of salted meat. It was incorporated in pulse puddings—some years ago Dr. James Pearse informed me that it was being used in this way by a family in the Lake District as part of the children's diet under the name of 'Easter May Giants'. An interesting and curious corruption of the linguistic hybrid Easter mangeant.

Other instances of profound physiological effects produced by extremely small amounts of chemical substances found in plants are afforded by the oestrogens present in clovers and the growth hormones. The recent discovery of the great potency of Gibberrelic acid to promote growth emphasizes how minute may be the stimulating cause. So too modern research has revealed the widespread occurrence in seeds of highly potent growth inhibitors. If some of these or, more probably perhaps, their synthetic analogues could be harnessed therapeutically the control of growth processes in man, whether malignant or benign, might not be chimerical.

Botanical science can perhaps make its contribution to the health problems confronting the human race which result from the disturbing augmentation of the world's populations, in various ways Substances present in species of *Lithospermum* have long been known to influence human fertility significantly and it is possibly not too much to hope that the investigation of these may, with the remarkable skill of the modern organic chemist in the field of synthesis, pave the way to the hoped-for oral contraceptive.

The role of the trace elements, which play so essential a part in the enzymic systems of the plant and animal body, is familiar, especially with regard to the elements copper and cobalt. What

is not so widely appreciated is that the earliest recognition of the necessity of these trace elements was established by plant physiologists stimulated by the effects produced by fortuitous observations. Maze's experiments demonstrating the necessity of zinc for plant life are reputed to have been inspired by the accidental use of a galvanized container for a spray that produced unexpected benefits, and the essential nature of molybdenum for plant life was brought to notice as a consequence of its presence as an impurity in mineral fertilizer.

The plant is normally the source whence the human body obtains the trace-elements it needs and, as we now know, the capacity of different kinds of plants to absorb and accumulate these metallic ions is extremely varied. The way in which Seaweeds accumulate iodine, which is familiar to most, is by no means unique. Species of Milk Vetch (*Astragalus*) exhibit a similar capacity for accumulating Selenium to an extent which can render the herbage on a seleniferous soil fatal to cattle and their flesh unfit for human food. So too the different grasses [illegible]

health in respect to radiation fall-out. It is recognized that if fall-out were to attain high proportions, radioactive strontium, because of its long half-life and because it can replace calcium in animal cells, would be one of the more serious risks. Moreover strontium is taken up readily by grass and cereals which tend to concentrate the element. On the other hand Lucern is much less prone to take up strontium than grass so that risk of absorption via milk is appreciably less if cows are fed upon the one rather than the other.

In times past when village communities were nourished upon plant and animal products almost wholly produced locally it is obvious that any trace-element deficiency inherent in the district's soil might have had a deleterious effect upon human health in a way that is most unlikely to occur under modern conditions when our food supplies are derived from so many different areas. In view of the high repute that many herbs had in the dietary of the past one may, however, suspect that perhaps

their differential capacity for accumulating trace-elements may have given some of them an importance in the past that does not now obtain. Unfortunately our knowledge of the trace-element contents of the foods we eat is woefully meagre.

One of our needs today is for a much wider survey of the quality of the food plants we consume. So little is as yet known concerning the new varieties that we raise and in the selection of which, yield-capacity and resistance to disease are the main criteria. I am not suggesting that our breeding efforts in relation to food values have been merely inflationary, or even wholly calorific, but it must be realized that until our assessment of new varieties is far more meticulous than at present we cannot be assured of the extent to which our efforts represent a real increase of our nutritional assets, or are largely illusions of increased bulk, that at the best may be mere calories and at the worst expensive water in an attractive guise. But I realize that perhaps the measurement of good food may be as liable to evade our existing techniques as the assessment of good health, the more so inasmuch as they are not absolutes but relativities.

A second approach to the problem is for the botanist to breed strains of crop plants not only capable of higher yields but with higher nutritive values. Thus the plant breeder has produced strains of Avocardo Pears which, it is claimed, have a protein value comparable to that of a good beef steak. But it is important to remember that plants, however great may be their potential fruitfulness, cannot build up substances that provide fuel for the human machine without utilizing the raw materials with which they are built up. This is so much a truism that it would seem scarcely worth uttering were it not that today there are not a few who appear to imagine that the increasing food supply that the human race requires can be provided without the use of artificial fertilizers. Not perhaps realizing it, they advocate a policy that really amounts to 'taking in one another's washing'. One such, whom I was able to convince that he could not continue to harvest his crops without putting back more

upon the familiar political standby of robbing Peter in order to pay Paul. The botanists may provide the more nutritional strains, but to realize their potentialities the nutritional status of the soil must be correspondingly enhanced.

It is I think instructive to remember that the usually very mixed character of the human diet has militated against the accumulation of empirical observations on the effects of prolonged consumption of food constituents that do not have an immediate and obvious action. Only the very restricted diet of the long sea voyages of the sixteenth century brought realization that scurvy was of nutritional origin and could be remedied by the use of fresh vegetables or lemon juice, though it was long before such observations led to the critical experiments that brought the recognition and isolation of the vitamin itself.

The much more restricted diet of domesticated animals produced empirical observations and subsequent experimentation that revealed not only important trace-element responses to copper and cobalt but also the presence in some pasture plants of oestrogenic substances, which, in certain strains of subterranean clover, are present in sufficient quantity to have very serious effects upon sheep grazed continuously upon it. The marked differences in the biological effects of different strains of one and the same species on stock suggests that this may well be true for human nutrition also and indicates the need for botanical and biochemical collaboration in respect of the less crude aspects of diet. Too often in food analyses, whether concerned with the grosser or more recondite constituents, no attention has been paid to the variety of the species investigated, still less to the circumstances of its culture. Here is an almost untilled field of human dietetics that might well yield a valuable harvest, whilst careful clinical observation of those who, from choice or from circumstance subsist on a very restricted diet, might well furnish important clues for future research to enlighten us in that *terra incognita* of the long-term effects of repetitive intake of plant products that are not manifestly harmful or beneficial when considered merely from the point of view of their short-term action.

One aspect of botany in relation to health to which I feel I

must refer is the alleged healthiness of plants and their suitability as human food, when grown with organic manures in contrast to the asserted unsuitability for human consumption of plants grown with mineral fertilizers. No informed person would wish to belittle the practical value of the colloidal material we term humus, which can abundantly retain upon its extensive internal surface both water molecules and nutrient ions, which alike supply the growing plant, whilst still permitting adequate soil aeration. But while recognizing all its virtues, we must not allow our imagination and sentiments unbridled licence so that we become muck-minded in every sense of that phrase.

When a plant grows in a soil containing organic material, whether from manure or compost, it gradually depletes the supplies of water and mineral nutrients that were retained upon the colloidal surfaces. Continued healthy growth of the plants demands that these supplies of water and nutrients should be alike renewed. The large demands made by modern high-yielding crop-plants which are repeatedly grown to maturity and removed, necessitate such replacement and the use of mineral fertilizers, properly applied, is in no real sense artificial but merely a replenishment of what is so continuously depleted. Because mineral fertilizers are so concentrated and far more easily handled than organic manures they can more readily be used to excess and abused, but I must most emphatically stress that there are no data that one could dignify as evidence to warrant the absurd assertion that plants grown with the use of mineral fertilizers are any less healthy than those grown with organic manures alone nor is there any real evidence to support the contention that such plants are in any degree less suitable for human food. The nostalgia for a countryside that is not converted into a food-factory is a sentiment with which most of us can sympathize, but this is the price we must pay for the unrestricted growth of population that is the greatest menace to civilization of today and holds the threat of global war as nothing else.

To turn to another quite different aspect, I would refer again to the relation of plants and animals to gaseous exchange. This

has been the subject of great oversimplification and misrepresentation. In its crudest form this is exemplified by the schoolboy howler that 'Animals take in Oxygen and give out Carbon dioxide whilst plants take in Carbon dioxide and give out Oxygen. If therefore a Rabbit and a Cabbage were placed together in a hermetically sealed chamber they would live for ever.'

The many queries I received from nurses and medical men when I was Director of the Royal Botanic Gardens, Kew, concerning the justification for the belief that it is harmless to have flowers in a patient's bedroom during the daytime but harmful at night, is sufficient evidence of the extent to which this oversimplification persists. It is true that green parts of plants, both foliage and stems, can carry on photosynthesis in adequate illumination and in this process they combine atmospheric carbon dioxide and release oxygen, but simultaneously the entire living plant is also consuming oxygen and releasing carbon dioxide in the process of respiration and this is often appreciably higher for flowers than for leaves. Even if this distinction be ignored, the light intensity in most bedrooms is so low compared with that in the open that the rate of photosynthesis that might be carried on by a vase of foliage would be below the rate of respiration. In technical language the leaves and flowers would almost certainly be working below their compensation point. So it is clear that the bedroom flowers if ever harmful in this way would be almost equally so both day and night. However, if the difference they could make to the gaseous composition of what the patient inhaled were in any degree significant the ventilation of the bedroom would I suggest be so inadequate that the presence of the flowers would merely anticipate but could scarcely hasten the inevitable wreath.

Thus far we have adumbrated a few instances where botany has been of direct benefit to the maintenance of health or the alleviation of disease but perhaps the benefits that medicine has derived from botanical studies have been equally, if not more, important from their contribution to the advancement of fundamental biological knowledge. We recall that it was botanist Gregor Mendel who laid the foundations of genetical science

and a plant physiologist Pfeffer who made the classical experiments that enabled the laws of osmotic phenomena to be formulated. So may I conclude by returning to the theme prompted by this course of lectures. Knowledge can be likened to a house with many windows through which we peer and endeavour to construct a picture of the whole from what we see. The narrow *specialist confines his attention to what can be seen through one* window only, but if he possesses the philosophical approach will realize the limitations of his perspective. If he is not of a contemplative mind he is more likely to add to our information than to our real knowledge. The dilettante is so interested in passing from window to window that he does not give himself time to apprehend the vision of any one. A true apprehension can only be approximated if we rectify our contemplative vision through one window by a reflective, though necessarily less intensive, examination of the prospect through others. Awareness of the unity of all knowledge is being more and more forced upon us and, though increasing specialization is a price we must pay for the ever-expanding bounds of science, the role of the general practitioner, both in your profession and others, assumes an augmenting importance, as a liaison officer, helping us to maintain a just perspective and to apprehend that the different viewpoints whilst presenting us with very diversified vistas are in fact, though so dissimilar, consistent aspects of the same fundamental unity.

II

The Control of Growth and the Cell Surface

M. ABERCROMBIE

It is asking for semantic trouble to use the phrase 'control of growth' without explaining one's meanings. Many biologists like to restrict the application of the word 'growth' to increase in mass, or more purist still, dry mass. I am going to be more comprehensive, and include in my subject-matter cell movement, which may bring about an increase in the space occupied by a population of cells. This is to a small extent what is commonly meant by growth of a malignant tumour, that is, in so far as its size depends on invasion. It is to a large extent what is meant by growth of an explanted fragment in tissue culture. Such a fragment may indeed lose mass, but by spreading, diluting itself so to speak in its surroundings, its volume of occupation increases. Purists are no doubt right in disliking this use of growth: it is liable, in the absence of precise analysis, to confuse disparate underlying processes But on the other hand the association, and in fact the causal relations, between spread by cell movement and increase in mass seem to be sufficiently close to justify treating them together, though not of course to justify confusing them.

The use of the world 'control' equally needs brief discussion. It is used I suppose to mean the specification of rate or amount of growth, the holding of growth to some quantitative pattern (which may of course be zero growth). Such a notion of control is, however, full of pitfalls unless all statements involving it are formulated with an idea of appropriate experimental testing in

mind. It has been all too easy to intone remarks like 'cancer is an escape from the normal control of growth exerted by an organism on its parts' and to feel one has been informative. But if a natural anarchist with no ingrained respect for control asks how you would propose to test this hypothesis, he will soon deflate the pronouncement to the mere fact that cancers grow where normal cells do not. It helps in keeping near the reality of experimental testing if one restricts analysis to the control of *changes* or *differences* in growth. As long as one is making a comparison, the classical requirement for an experiment, a control group, is part of the picture; this second sense of 'control' has profound practical affiliations with the first.

Growth of Organs

A great deal of work on the control of growth has been concerned with the size of embryonic and larval organs. Such work forms an essential background to any analysis into cellular events A striking feature of the developing vertebrate is that with increasing age there is a continuous decline of specific growth rate (that is, of growth per unit of material already formed). Though subject to local fluctuations, this decline is general and fairly consistent amongst the different organs of the body. It is a change in growth, and therefore discussion of its control can be conducted in experimental terms, which makes it a useful example to consider. A natural presumption is that it represents a steady systemic change in some component of the internal environment of the embryo. If this is so, then an organ transplanted from, say, an older donor to a younger host should show the higher growth rate of the younger stage. The experiment has often been done in Amphibia, with limbs and eyes transplanted from older to younger and vice versa (see Twitty and Elliott, 1934; Twitty, 1940). It has turned out that there is no systemic decline in growth-promoting power of the internal environment capable of influencing these organs. An older graft in a younger host is not speeded up; apparently indeed it grows rather more slowly than it would if transplanted to an embryo of its own age, it is overhauled in size by the corresponding host organ, losing its age advantage. Conversely

a younger graft in an older host is certainly not slowed down, and apparently grows rather faster than it would on a host of its own age: it overhauls the corresponding host organ. In their new situation transplants from young to old or from old to young, judging by the published growth curves, before long show the gradual decline in specific growth rate so characteristic of vertebrate development.

Since there is no general decline of the growth-promoting power of the internal environment with age, but on the contrary an increase, Twitty (1940) drew the conclusion that, in limbs and eyes at least, there is within each organ an autonomous decline in 'assimilative ability' which is independent of what the rest of the embryo is doing, and is sufficiently pronounced to offset the increased growth promotion by systemic components. If one were to extrapolate from the behaviour of limbs and eyes, one would conclude that the developmental decline in specific growth rate is general and consistent because all organs have built into them a similar independent pattern of growth control. There is some hint as to the mechanism of this autonomous control within each organ. It does not seem to work simply by tapering off growth with advancing age. According to Twitty, although younger transplanted eyes in older hosts catch up the host eyes in size, they do not surpass them as might be expected if growth rate depended simply on age and systemic conditions; having reached the same size, they grow thereafter at the same rate as the host eyes, in spite of the difference in fertilization age. These observations perhaps need putting on a firmer basis (see Handford, 1948); but if they prove to be acceptable it seems that the control of the decline in growth rate involves a tapering off of growth with advancing *size* of the organ, not age, or at least with something normally closely correlated with size; in effect some feed-back from the products of growth.

The idea of feed-back from the products of growth brings us to another hypothesis, which accounts for the sort of experimental result we have considered, but does not allow the local autonomy asserted by Twitty's hypothesis. This second hypothesis is a feed-back system proposed by Weiss (1955; Weiss and

Cavanau, 1957), which has proved very attractive to many workers. He suggests that each cell type of a growing animal produces into the extracellular space of the whole organism its [illegible]

[illegible]

[illegible] tinued production of such anti-templates leads to a declining growth rate. This mechanism puts the growth of any tissue under the influence of the size of the rest of the organism, because the concentration of anti-templates and therefore their effectiveness depends not only on their output by the tissue but also on the volume in which they are dispersed. In terms of the transplantations of limbs and eyes we have been considering, when we put, say, a young eye on an old host we would be putting a small producer of anti-templates into a large receptacle for anti-templates. Their concentration should fall and promote the growth of the transplant. By the time the size of the grafted eye becomes appropriate to the size of the host, the anti-template concentration and consequently the growth rate will have risen to the normal for that stage, or at any rate roughly so. (In the sophisticated version of the theory developed by Weiss and Cavanau in 1957, anti-template concentration at any moment does not depend simply on volume of receptacle and on total mass of producing organ at that moment, so precise regulation of transplant size to host size is not to be expected.)

Since Weiss' hypothesis involves correlation of tissues via a systemic pathway, it is tested by any procedure on an embryo which changes the ratio of the amount of the organ in question to the volume of the whole organism. Such a change should produce a compensatory change in growth of the organ, as indeed it does in the experiments of Twitty's I have quoted. It seems, nevertheless, unlikely that Weiss' hypothesis can apply to limbs and eyes because other experiments that alter the ratio do not apparently alter growth of these organs. Compensatory hypertrophy of a limb or its parts has not as far as I know been observed, nor depression of the growth of the homologous limb when an extra limb is grafted on to an animal, nor, when grafts are made between two closely related species one of which

grows much bigger than the other, does there seem to be any effect on growth rate of limbs or eyes through the change in the size of the whole organism (e.g. Church, 1956). It seems that Twitty's hypothesis where the feed-back, if any, is localized to the organ, is preferable at least for eyes and limbs, and presumably for other parts of the musculo-skeletal system.

Nevertheless, it is highly probable that there are several important organs that grow by a mechanism at least formally similar to that proposed by Weiss. These are the organs that undergo compensatory hypertrophy in the adult, such as liver and kidney. During the later part of the life history it seems that the size of such an organ is controlled by the blood level of some regulatory substance produced by the organ itself (see Abercrombie, 1957), though whether such a substance acts directly as an anti-template is an entirely open question. We are perhaps not yet on entirely safe ground in supposing that when com- [illegible]

hypertrophy (e.g. Fox, 1956). Such compensatory mechanisms in a growing animal will adjust size of their particular organs to that of the rest of the body. If the rest of the body shows a decline in its growth rate, they will too.

The hypotheses of Twitty and of Weiss of course far from exhaust the suggestions that have been put forward about control of growth. One further idea, which might apply to the decline of growth rate during development, must suffice. Crile (1958) has suggested that a feed-back mechanism exists of rather greater complexity than that of Weiss' hypothesis; each cell type he supposes produces a signal substance which inhibits the production from some other site in the body of a stimulating substance specific to that cell type. Such a mechanism will induce compensatory hypertrophy if the number of cells producing the signal substance is experimentally reduced; and where we find that compensatory hypertrophy does not occur we can rule out such a mechanism, except as a phenomenon so localized that both signalling cells and stimulating cells are simultaneously reduced in number by the operation. There is

no reason yet for supposing that this mechanism is widespread in an organism. But as with Weiss' hypothesis, it is highly probable that the growth of *some* embryonic organs is influenced in this way. Some of the endocrine glands that are linked to the pituitary, notably the thyroid (which formed the model for Crile's theory), work like this in the adult, and the thyroid-pituitary axis is established early in development (see for instance Jost, 1954). If the consumption of thyroid hormone varies with the size of the whole organism, so will the size of the thyroid.

In the particular example of a growth change which we have been considering, the developmental decline in the specific growth rate, we seem to find different mechanisms at work for different major organs. Organs like limb and eye, which appear so independent in their growth rate changes, need not of course be a mosaic in their internal organization. Indeed for the eye it is known that there are internal interactions: transplantations show that an overlarge lens in an oversmall eye-cup grows more slowly itself, and makes the eye-cup grow faster (Harrison, 1929; Ballard, 1939). Nor does the existence of a systemic self-regulatory system like that of liver or thyroid imply that there are no interactions between the various subdivisions of the total cell population of the organ. For instance, in the adult liver undergoing compensatory hypertrophy after part of it has been removed, all the populations of different cell types grow more or less appropriately, and the most probable explanation is that the parenchymal cells are setting the pace and in various ways influencing the other cells (Abercrombie and Harkness, 1951; and see Santler, 1957, for a similar situation in the thyroid stimulated by thyrotrophic hormone). The same may be true during development.

Enough has been said about the particular problem of the declining growth rate to support the point of view that patterns of ontogenetic growth are specified by a variety of mechanisms, varying from organ to organ and tissue to tissue. This complex situation must be worked out piece by piece. The tendency which has so often shown itself to promote just one kind of control mechanism to pre-eminence in the embryo is probably far too superficial an approach to the problems. The same

picture of variety of mechanism emerges when one considers various instances of proliferation in the adult (Abercrombie, 1957; Swann, 1958).

The Control of Locomotion

It seems, however, that the growth process itself, unlike the mechanisms which control its rate and amount, may have an unexpectedly far-reaching unitary character. It obviously consists of protoplasmic synthesis plus mitosis, a cellular response which may be considered to be induced by the controlling mechanisms in somewhat the same sense that differentiation is induced (Swann, 1958). It comprises also locomotory activity by the individual cells (Abercrombie, 1957; the formation of intercellular substances is also probably part of the syndrome but will not be considered here). This is the justification for including cell movement under the general heading of growth. The complexity of the growth response is understandable because its outcome is the production of organized, complex tissue, which cannot be formed by mere multiplication of cells *in situ*, but requires also that new cells move into the appropriate arrangements.

The locomotory apparatus of cells, that is, whatever it is that enables them to crawl, individually or in coherent groups, on a suitable solid substratum, seems to be in an activated state wherever growth involving mitosis is going on. This activated state may be assessed by explanting fragments in tissue culture. The more the activation, the greater the number of cells that will emigrate from the fragment in a given time. Locomotory activation thus tested is in general correlated with mitotic activity. In the embryo both are high at early stages, and both decline together as development proceeds (see Medawar, 1940). In episodes of induced proliferation in the adult, such as liver regeneration or thyroid hypertrophy, both change substantially together (see Abercrombie, 1957). Locomotory activity tested in this way is, however, not necessarily a direct measure of the amount of locomotion that occurs *in vivo* in the proliferating tissue. It probably represents a general excitation of the locomotory mechanism of the cell, which manifests itself in the form of

mass emigration in the rather special situation in tissue culture. In the organism, one would expect there to be detailed mechanisms for directing the movement of individual cells and for stopping it at the right places. The study of cell behaviour, which has a long history behind it, has in fact shown the existence of such mechanisms, and this brings up the first instance of an important role of the cell surface in the control of growth processes.

As a result of investigating the movement of fibroblasts in tissue culture, it has appeared that a significant part in its control is played by the mutual relations of the cells; one reaction, termed contact inhibition, seems particularly important as a mechanism of precise adjustment of movement (Abercrombie and Heaysman, 1954; Abercrombie, 1957). This reaction occurs when a moving fibroblast comes into contact with another fibroblast, and as a result stops moving forward. It may soon, however, start moving again in a new direction. This effect acts as a potent orientating agent for fibroblast movement in tissue culture, since it produces a strong tendency for a population of active fibroblasts to move into any neighbouring cell-free space (Abercrombie and Heaysman, 1954). Contact inhibition can also of course stop locomotion, since a fibroblast so situated that in whichever direction it starts to move it makes contact with another will be reduced to mere oscillations. The relatively stable adhesions that fibroblasts form with each other, [illegible] result in contact inhibition, movement within a population may to a limited extent be possible.

It seems that behaviour similar to that we have called contact inhibition occurs between sheets of epithelia. It was described, long before we analysed it in fibroblasts, by Howes (1943) who investigated epidermal movement in mammalian skin wounds, and concluded that movement continues until one sheet runs into another. Chiakulas (1952) studied in amphibia the mutual behaviour of a variety of epithelia transplanted to the skin surface, and showed that some kinds inhibited each other's

movement on contact while others did not. Hence it seems that at least the stopping role of contact inhibition can be discerned amongst epithelia.

If we now consider the nature of contact inhibition, the only point that is pretty clear is that it pre-eminently concerns the cell surface. Contact inhibition only comes into action when contact is actually made: we have never been able to detect any action at a distance, so that mutual interference with diffusion away from the cell surface into the medium seems unlikely to be the mechanism. Since the inhibition occurs not only when two similar cells but two similar parts of such cells touch, it seems hardly likely that one cell can give the other a substance it has has not got already: so it surely does not involve an exchange of inhibitors. Some kind of direct juxtaposition of the cell surfaces seems to be the relevant inhibitory situation.

As well as contact inhibition there is another potent influence on cell movement, the contact guidance of Weiss (1941 and earlier). An orientated structure of the substrate imposes orientation on the cells: movement along the grain of the substrate is strongly promoted, movement across it is difficult. The importance of the cell surface in this reaction is obvious. Unidirectional movement is not specified in this way for that, cells have to be given some other clue, which in culture is normally provided by contact inhibition.

Reverting now to the test for locomotory activation by explanting fragments, it clearly provides in the vacant regions of the culture medium cell-free space into which cells subject to contact inhibition will tend to move. Nevertheless, as we have seen, cells do not automatically take advantage of this space, but only to the extent that the tissue explanted is already in a proliferative state. It appears that the sort of activation associated with the induction of mitoses *and* freedom from contact inhibition are both required for successful movement. It should be emphasized that tissue culture has so far provided the only way that these problems can be investigated and extrapolation to behaviour *in vivo* requires the bridging of a considerable gap.

A proliferating tissue differs from a resting tissue in the state of the locomotory mechanism of its cells. The difference may be

in the cell organization by which energy is diverted to particular functions, or it may even be that the structural materials required for locomotion are present in the former and missing in the latter. But it can be argued that the cell surface may also be involved in the state of activation to an important degree. The attachment of a cell to surrounding structures may impose a significant limitation to its movement. We found evidence that close attachment to a substrate may have a profound effect in limiting ability to move in the admittedly rather special case of the attachment of Schwann cells to nerve fibres (Abercrombie, Johnson and Thomas, 1949). If attachment to substrate is important, the process of activation may involve a release from such inhibitory adhesions. It is interesting that trypsin, which may be supposed to do just that, distinctly increases locomotory activation as tested by explanting treated fragments (Simms and Stillman, 1937).

The Control of Mitosis

There is then evidence that, for some cell types at least, there is a fine-grain control of locomotion effected by contact reactions, superimposed on the large-scale controls regulating proliferation which we considered at the beginning and which themselves may also involve contact reactions between cells and substrate. Can we implicate the cell surface to any important extent in the control of mitosis, the second main component of the proliferative syndrome? The hypothesis has taken clear form with the suggestion made by Dr. E. J. Ambrose, and independently by Swann (1958), that mutual contact of cells may influence their mitosis through an effect on uptake of nutrients. As Ambrose and I have observed (1958) a striking effect of mutual contact between fibroblasts is the abolition of membrane movement and hence of the associated uptake of fluid droplets (pinocytosis) in [illegible] individual mitoses could conceivably be linked to mutual contact in this way, in the same way that locomotion is controlled; though no evidence of this has yet been detected in cultures.

And the general control of amount of mitosis in a cell population which occurs through the various mechanisms discussed earlier in terms of whole organs, could also take place through a mobilization of the cell surface, by detachment from substrates or in some other way that influences intake of substances. It is conceivable, though far-fetched at present, that the proliferative syndrome in both its mitotic and locomotory aspects is mediated through changes in the cell surface, both in general level of activity and in precise location of individual acts of mitosis and movement.

Malignancy

Finally I want to consider the relation of malignant growth to the normal growth I have been discussing. A population of cells undergoing malignant growth manifests the usual combination of mitosis and of locomotory activity (as tested by explantation). The proliferative syndrome is apparently permanently switched on. But this well-known relative autonomy of proliferation is not apparently the only malignant property. So far as our limited tests go (Abercrombie and Heaysman, 1954; Abercrombie, [illegible]

culture the stopping of movement when they collide with normal fibroblasts that normal fibroblasts do when they meet each other. This is not unexpected if we grant the possibility that contact inhibition *in vivo* holds even activated fibroblasts in place and consequently does not permit invasion. The failure of contact inhibition would then make invasion possible for sarcoma cells. This difference in malignant cells, if for the sake of argument it may be so generalized, points strongly to a peculiarity in their surface properties. Much other evidence points the same way for instance Coman's classical demonstration of reduced adhesiveness of carcinoma cells to each other (see Coman, 1953) and the work of Ambrose and his colleagues showing a change in electrophoretic mobility of malignant cells (Ambrose, James and Lowick, 1956).

In malignancy we can therefore suggest that there are two deviations from normal within the field I have been discussing:

permanent proliferative activation and loss of contact inhibition by normal cells, a change in both coarse-grain and fine-grain controls. It is not of course ruled out that these may be independent changes which have coincided in a cell and hence in its offspring. The statistics of cancer incidence allow one to postulate independent changes in carcinogenesis. But I have already indicated the possibility that the various normal controls of the proliferative syndrome may involve the final common pathway of surface change. It is therefore possible that these two changes from normal in malignant cells may both be brought about by a transmissible change in surface properties of the cell.

Conclusion

This discussion has clearly been slanted to emphasize a point of view, that the control of growth may be very closely linked with the properties of the cell surface. One must beware, however, of giving undue primacy to one component of the cell, and of merely enthroning the surface in the place of the DNA. If the surface is important in growth processes, it is probably only as part of a complex mechanism involving the rest of the cell. I do not want to do more than conclude that the cell surface should not be neglected in the study of growth.

REFERENCES

ABERCROMBIE, M (1957). *Symp Soc exp. Biol.* **11**, 235
ABERCROMBIE, M and AMBROSE, E. J (1958). *Exp Cell Res.* **15**, 332.
[illegible]
Exp. Cell Res. **13**, 276.
ABERCROMBIE, M., JOHNSON, M. L and THOMAS, G A. (1949) *Proc. roy. Soc.* B, **136**, 448.
AMBROSE, E J, JAMES, A M. and LOWICK, J H B (1956). *Nature, Lond* **177**, 576.
BALLARD, W B (1939). *J. exp Zool* **81**, 261.
CHIAKULAS, J L. (1952). *J. exp Zool.* **121**, 383
CHURCH, G (1956) *J. Morphol* **98**, 405.
COMAN, D. R. (1953). *Cancer Res* **13**, 377.

Crile, G. (1958). *J. nat. Cancer Inst.* **20**, 229.
[illegible]
[illegible] 103.
Swann, M. M. (1958). *Cancer Res* **18**, 1118.
Twitty, V. C. (1940). *Growth*, (suppl.) 2nd Symposium on Development and Growth, 109.
[illegible] 68, 247.
[illegible] G. Butler, 195.
Weiss, P and Cavanau, J L. (1957). *J. gen. Physiol.* **41**, 1.

III

Electron Microscopy and the Living Cell

E. H. MERCER

ELECTRON microscopy is a novel practice and it still stands in need of some apologia, particularly when it makes bold to have opinions concerning structures in the living cell. In his specific capacity, the electron microscopist never looks at a living cell, nor is there any immediate prospect of his doing so. His material is, to say the worst about it at the very beginning, the fixed and dried-up residue of a cell which is fried *in vacuo* by a beam of electrons. Nevertheless, he hopes, along with the light microscopist, the biochemist and others, to make a contribution towards the understanding of the living cell and, being the inheritor of a long tradition in which the reproachful word 'artifact' regularly crops up, he has had to become more sophisticated in his approach than his predecessors. What I wish to describe here is something of the tactics of electron microscopy in biology and to show that, properly appraised, it has much the same sort of validity as other methods of study. Like them, in isolation it can be deceptive, but when integrated with other [illegible] directly on the [illegible] ultimate goal [illegible] rms.

The electron microscopist's information comes to him in the form of an image on a photographic plate. He knows that this is not the image of something living; it is in fact the image of a reaction product formed between an originally living structure and certain chemicals. He has in effect carried out a mortal experiment on a living object and the photograph records an

aspect of the results of that experiment. The problem is to deduce something of the nature of the original object from the image and to correlate this information with information drawn from other sources—paramount among these being the observation of the cell in life. Boiled down to laboratory tactics, this means in practice that we need to know all we can about the cell before we start, and that the experiment must be controlled as long as possible, during the life of the cell and during processing, by observation in the light microscope. This may prove difficult in practice, for example, owing to the location of the cell in a tissue, and of course it is limited in resolution. The real arguments after all are going to concern what cannot be seen in this way. Nevertheless it remains that, at the very least, a sufficient amount of work must be done to show that nothing deleterious has occurred as far as can be seen using light. We cannot say that we aim at perfect preservation of the cell, for this has no meaning unless we imply instantaneous freezing at $-273°$ C. The aim is a *useful* preparation, which is one that enables us to obtain the information we want even at the cost of losing other information we do not need to obtain in that particular experiment. A moment's thought will make it clear that in principle all experiments have the same limitation; we cannot observe the external world without disturbing it, and we trade the disturbance we produce against the information received (Brillouin, 1956, 1959).

It is necessary to make these very general remarks in order to introduce a point of view and a vocabulary. Discussions with hostile critics tend rapidly to become bogged down in disputes as to the validity of one method as against another. Whereas the real problem is how to extract from each the information it can give and to integrate it all into a coherent picture. Electron microscopy is not alone in posing this problem—the physicists have it too—but the new microscopy has certainly brought biologists face to face with it. Fortunately the new developments in information theory point the way to a solution (Brillouin, 1959). We shall try to adopt the view that experiments are carried out to obtain information about the external world in order to adapt ourselves to it. We do not 'see, hear or feel' this

external world but from 'information received' we built a picture of it (Russell, 1958).

Techniques of Preparation

There is no need here to go into the details of fixation, embedding, etc., but sufficient must be said to indicate the nature of the problem and the steps taken to assess the changes it produces. The procedure follows the sequence of fixation, dehydration, embedding and sectioning already familiar from light histology (Fig. 1). Changes in structure are produced by everything we

Operation	*State of cell*
1 killing	dying
2 fixation	dead and stabilized
3 dehydration	dehydrated
4 embedding	embedded
5 sectioning	cut into thin sections

Fig. 1. Sequence of operations for preparing biological material for electron microscopic examination in thin sections

do but these are only dangerous when we do not admit they have happened and thus cannot make allowances for them. The fixative also kills the cell. Some workers, hoping to coax it into the grave without harming it, at first advocated a 'gentle fixation' or 'slow killing'. This is probably misplaced humanitarianism. Organisms suffer less when killed quickly and translated into cytological terms this means: when a poison enters the cell, the homostatic mechanisms begin at first to make adjustments within the normal range. Then, as various elements fail, the adjustments take abnormal forms. Since, at least to begin with, we do not wish to study these abnormal forms, the aim should be to knock out everything in one step. This is the reason for using very small blocks (<1 mm.^{3}) or single cells to facilitate penetration and to maintain a high concentration of fixative; it is probably one of the reasons for the effectiveness of osmium tetroxide, the fixative *par excellence* of electron microscopy (Palade, 1952a). Being soluble both in aqueous and non-aqueous phases it penetrates each with equal facility.

During fixation some components are stabilized, others not and some are removed. We do not know enough about this step.

It is one of our blind spots since the chemistry of osmium tetroxide is poorly understood and alternative procedures are still being developed. When at first the procedures found valuable for light microscopy were tried, it was immediately obvious that these grossly deranged the fine structure. This might well have been expected since one object of fixation is to produce something to see and, for the light microscope, this meant the precipitation and aggregation of fine structure to produce a texture of the order of at least 0·2–0·3μ in diameter. Fixatives such as osmium tetroxide, which were known to produce a life-like preservation of the cell (Strangeways and Canti, 1927), were for this (among other reasons) not greatly favoured. No doubt other suitable fixatives exist. Buffered neutral formaldehyde has its uses and, unexpectedly, potassium permanganate was found to be an excellent selective fixative (Luft, 1956).

Washing can have disastrous effects when unstabilized components swell; dehydration, on the contrary, has given less trouble, since the physical strains in the small blocks are well tolerated. The embedding process has occasioned more difficulties because, under some conditions, the commonly used polymethacrylates can cause swelling during polymerization. This has been largely overcome by the introduction of cross-linked condensation polymers (Glauert *et al.*, 1956; Kellenberger *et al.*, 1956).

The changes produced by each of these steps can be followed by means of light microscopy if material suitable for constant observation is chosen. Tissue cultured cells are admirable for this purpose (Porter, 1957), as also are large protozoa. My colleague, M. S. C. Birbeck, and I have studied the preservation of the large amoeba (*A. proteus*) and this work may be described as an example (Birbeck and Mercer, 1957a and b). Phase microscopy and dark-ground methods were found most convenient. As far as could be judged by eye, preservation by the buffered osmium tetroxide was perfect and was not impaired by washing and dehydration. Prolonged exposure to absolute ethanol (24 hours) led to increased light scattering in the cytoplasm and in the plasma membrane. The most damaging step was found to be the embedding in methacrylate. During polymerization

localized swellings were obvious and were later confirmed electron microscopically. These seemed to be due to uneven polymerization caused by local catalytic effects. When a small volume polymerizes in advance of its neighbours, monomer from these penetrates and causes the polymer to swell, thus distorting the embedded structures. Damage due to this cause was eliminated by adopting the entirely different type of embedding medium, the Araldite mixture introduced by Glauert, Rogers and Glauert (1956). Our experiments on amoebae (1957b) showed that its use led to no detectable swelling damage and the electron microscopic image was excellent.

Several other studies of this sort have been made and it has been established that for most materials the preservation is perfect as judged by light microscopy. A much more difficult question arises in assessing the validity of structures beyond the resolving power of the light microscope. How are we to know whether these existed in life? Obviously we must make appeal to alternate methods of structure determination and to criteria of a different kind.

(a) Independent techniques

When a structure contains crystalline or quasi-crystalline structures, their presence may also be detected by X-ray diffraction, sometimes in the living material. Examples are found among the many tissues having a fibrous texture, genuine crystals of proteins or virus particles, regularly laminated formations, such as the myelin sheath of mammalian nerves. Many of these formations may also be birefringent and thus be detectable by means of polarized light microscopy (Schmidt, 1937) which has an advantage over X-ray diffraction in being able to detect more minute organized structures in a cell.

A close quantitative agreement between the measurement of a periodic structure by means of X-ray diffraction and by microscopy gives confidence not only that the periodic structure [illegible] res,

sur-

(b) Internal evidence

Judgements based on the internal evidence of the micrograph are somewhat more subjective; nevertheless, as experience grows, we gain confidence in our ability to recognize a good preparation, since we are able to compare it with many others, some of which at least have had their validity established by other means. Various forms of damage or distortion have already been characterized and are now routinely recognized.

(c) Processing of artificial models

The behaviour of biological materials during preparative procedures may be evaluated by preparing artificial models from purified substances of biological origin and submitting these to routine fixing, embedding, etc. These models may be made to resemble more or less closely structures found in cells and they may be more open to alternative determination than their cellular analogues. We can in this way estimate the probable reactivity of various constituents with fixatives and stains and possibly also their modifications.

Such experiments help to identify the cellular components which react sufficiently with the fixative to display an increased density, i.e., to be 'stained'. The reaction of cell constituents with osmium cannot yet be placed on a predictive basis. Everyone seems to know that unsaturated lipids react strongly; in fact the test-tube experiments of Bahr (1954) revealed reactivity with a wide range of purified substances and, as a consequence, little selective staining.

(d) Compatibility with other evidence

Lastly there is the compatibility of the findings with all the other evidence bearing on the behaviour of the living cell. This is the integrative step, the attempt to bring all our information together to form a unified conception of the cell which will enable us to understand and predict its behaviour. In case the operation of this test is not clear, let me take a simple example. The mitochondrion is to the light microscopist a small object about a μ in diameter, possessing some fairly well-defined staining reactions, found in great numbers in the cytoplasm of almost

localized swellings were obvious and were later confirmed electron microscopically. These seemed to be due to uneven polymerization caused by local catalytic effects. When a small volume polymerizes in advance of its neighbours, monomer from these penetrates and causes the polymer to swell, thus distorting the embedded structures. Damage due to this cause was eliminated by adopting the entirely different type of embedding medium, the Araldite mixture introduced by Glauert, Rogers and Glauert (1956). Our experiments on amoebae (1957b) showed that its use led to no detectable swelling damage and the electron microscopic image was excellent.

Several other studies of this sort have been made and it has been established that for most materials the preservation is perfect as judged by light microscopy. A much more difficult question arises in assessing the validity of structures beyond the resolving power of the light microscope. How are we to know whether these existed in life? Obviously we must make appeal to alternate methods of structure determination and to criteria of a different kind.

(a) Independent techniques

When a structure contains crystalline or quasi-crystalline structures, their presence may also be detected by X-ray diffraction, sometimes in the living material. Examples are found among the many tissues having a fibrous texture, genuine crystals of proteins or virus particles, regularly laminated formations, such as the myelin sheath of mammalian nerves. Many of these formations may also be birefringent and thus be detectable by means of polarized light microscopy (Schmidt, 1937) which has an advantage over X-ray diffraction in being able to detect more minute organized structures in a cell.

A close quantitative agreement between the measurement of a periodic structure by means of X-ray diffraction and by microscopy gives confidence not only that the periodic structure

(b) Internal evidence

Judgements based on the internal evidence of the micrograph are somewhat more subjective; nevertheless, as experience grows, we gain confidence in our ability to recognize a good preparation, since we are able to compare it with many others, some of which at least have had their validity established by other means. Various forms of damage or distortion have already been characterized and are now routinely recognized.

(c) Processing of artificial models

The behaviour of biological materials during preparative procedures may be evaluated by preparing artificial models from purified substances of biological origin and submitting these to routine fixing, embedding, etc. These models may be made to resemble more or less closely structures found in cells and they may be more open to alternative determination than their cellular analogues. We can in this way estimate the probable reactivity of various constituents with fixatives and stains and possibly also their modifications.

Such experiments help to identify the cellular components which react sufficiently with the fixative to display an increased density, i.e., to be 'stained'. The reaction of cell constituents with osmium cannot yet be placed on a predictive basis. Everyone seems to know that unsaturated lipids react strongly; in fact the test-tube experiments of Bahr (1954) revealed reactivity with a wide range of purified substances and, as a consequence, little selective staining.

(d) Compatibility with other evidence

Lastly there is the compatibility of the findings with all the other evidence bearing on the behaviour of the living cell. This is the integrative step, the attempt to bring all our information together to form a unified conception of the cell which will enable us to understand and predict its behaviour. In case the operation of this test is not clear, let me take a simple example. The mitochondrion is to the light microscopist a small object about a μ in diameter, possessing some fairly well-defined staining reactions, found in great numbers in the cytoplasm of almost

all cells. To the biochemist it is a small container in which is found a number of enzymes, some closely associated with its surfaces, some more free in the sense that they enter solution when the container is broken. When electron micrographs of mitochondria were produced showing a double-walled enclosure containing numerous internal sub-compartments separated by membranes (Palade, 1952b), there was little hesitation in accepting the finding since it satisfied so adequately the biochemists' demands (Plate I, Fig. 2[1]).

Membranes and Organelles

The first step in correlating light and electron microscopy is to recognize the electron microscopic images of familiar objects, such as nuclei, mitochondria, inclusions, etc. This presents little difficulty since at low magnifications the electron microscopic images are essentially similar to their light counterparts. Much of this work is now familiar. It will perhaps be more interesting to turn to the description of certain structures, now coming to occupy an important place in cytology, which were scarcely suspected to exist, or were much misunderstood, until examined electron microscopically. What brings these together in a structural classification is that they are all constructed from membranes. The revelation of the extent of membrane systems in the cytoplasm is one of the most important contributions of the new microscopy, one again in excellent accord with the anticipations of biochemists who long predicted the existence of some form of cytoskeleton or framework (Peters, 1937).

Cytoplasmic membranes are particularly well developed in glandular cells which are secreting protein and their typical appearance is shown in Plate I, Figure 4. These consist of membranes covered with small dense particles (diameter 100–200 A) forming usually stacks of flattened sacs but capable of assuming a great variety of other forms leading to an immense profusion of confusing profiles when seen in section. These 'rough-surfaced membranes' were discovered almost as soon as sectioning was attempted in France, the U.S.A. and Sweden (Hagenau, 1958) and most observers immediately guessed that they were involved

[1] The plates referred to in this lecture will be found between pages 40–1.

in protein synthesis. It has required a great deal of cross-comparison of the results from living cells, fixed cells, electron microscopy, light microscopy, physiology and biochemistry to prove this connection. Now it is well established that the ultraviolet absorbing regions of the living cell and the basophilic regions, already inculpated in protein synthesis, are the regions containing the particle-studded membranes (Palade, 1955; Palade and Siekevitz, 1956). The break-up of these membranes produces microsomes (Plate I, Fig. 3) which are immediately recognized as smaller versions of the particle-studded membranes. We shall return to this characteristic behaviour of membranes later. We owe to Palade (1955) the idea that the membranes themselves are not basophilic; that this property resides in the small dense particles which contain the ribonucleic acid more directly associated with synthesis.

Such particle-studded membrane systems are only found in *glandular* or *secreting* cells. The other great class of synthetic cells, which may be called *retaining* cells, since they accumulate their protein and do not extrude it, have large numbers of dense particles but these are not attached to membranes (Birbeck and Mercer, 1957c) (Plate II, Fig. 5). It would seem likely then that the membrane system is concerned with the handling of metabolites, the orderly transport of material to and from the sites of synthesis or assembly.

Another group of organelles is formed from smooth-surfaced membranes and among these the Golgi 'apparatus' is conspicuous (Plate II, Fig. 6). The volume of the cell considered by light microscopists to be the Golgi region is usually found in electron micrographs to contain a curious cluster of smooth membranes usually forming a characteristic pattern of flattened sacs and smaller vacuoles (Fig. 7a).

It is the opinion of the influential Rockefeller group that all these membranes form an interconnected whole, parts of a single system which divides the internal volume of the cytoplasm into two parts. They refer to it as the *endoplasmic reticulum*. The term does emphasize the common membranous framework of many cell organelles but is not universally accepted (Hagenau, 1958). Certainly 'endoplasmic' is not accurate. Further, not

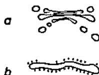


FIG. 7. Organelles based on membranes: (*a*) the cluster of vesicles and flattened sacs formed from smooth-surfaced membranes (γ-cytomembranes) found in the Golgi region (see Plate II, Fig. 6); (*b*) flattened sacs of particle-covered membranes (α-cytomembranes) as found in the basophilic regions of the cell (see Plate I, Fig. 4).

everyone agrees that all the surfaces are interconnected. Such a picture (Fig. 8) seems too static. Nevertheless from time to time lucky sections have revealed all the postulated connections, leaving little doubt of a transient continuity at least and offering a proof of the identity of all the membranes.

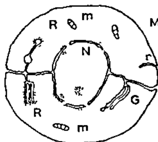


FIG. 8 A hypothetical conception of the possible interrelation of all the membranous organelles of the cell M, cell membrane, γ, cytomembranes formed by infolding of the plasma membrane; R, basophilic membranes (α-cytomembranes); G, Golgi type membranes (β); m, mitochondria and N, double nuclear membrane.

Needless to say, there are other systems of nomenclature. Sjöstrand (1953) believes that the membranes always occur in pairs and names them as follows:

α-cytomembranes = rough-surfaced basophilic membranes
β-cytomembranes = smooth-surfaced paired membranes, arising from invaginations of the plasma membrane
γ-cytomembranes = smooth-surfaced pairs

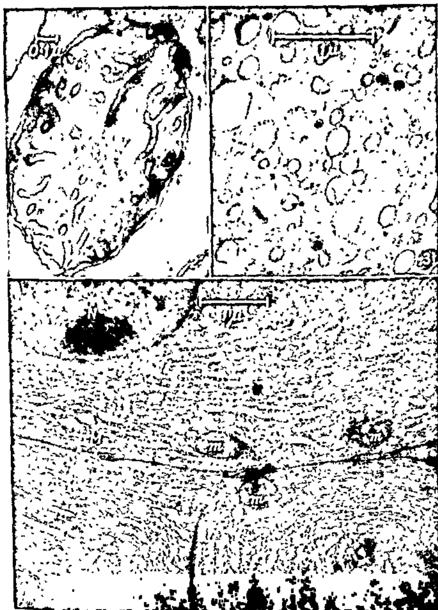


FIG 2 Cross-section of a mitochondrion from *Amoeba proteus* showing the double external membrane and the tubular invaginations of the inner surface Osmium fixation

FIG 3 Cross-section of a microsomal pellet separated by fractional centrifugation from homogenized rat liver Typically small vesicles covered with small dense (RNA) particles (see Fig 4).

FIG 4 Section of rat pancreas, a typical glandular cell showing junction of three cells, N ucleus, *m*, mitochondria, M, cell membranes The cytoplasm is tightly pac article covered membranes (α-cytomembranes or reticulum— roken up these give rise to microsomes (Fig 3) Osmium fixation lly lent by Mr M, S C Birbeck

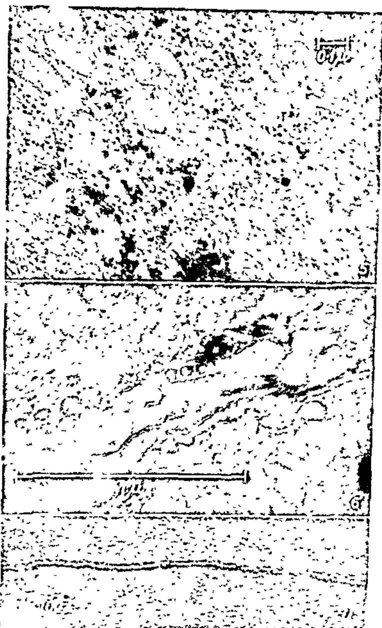


FIG 5 Section of a cell in a human hair follicle illustrating the typical cytology of a retaining cell, i e a cell synthesizing protein but not secreting it to be con-
The same dense
t associated with
)smium fixation

FIG 6 *A cluster of agranular vesicles and membranes located in the Golgi region of a rat liver cell* (See Fig 7*a*) Micrograph kindly lent by Mr M S C Birbeck

FIG 11 Cross-section of the plasma membrane of *Amoeba proteus* (See Fig 13)



FIG. 12 Section of a food vacuole in the absorption stage in an amoeba Undigested residues and small vacuoles budding-off the larger (Fig 9c) are to be seen Fixed in osmium tetroxide

FIG. 15 Section of a crumpled bundle of a layer of egg albumen spread on water (see text for explanation) Fixed in permanganate



FIG 5 Section of a cell in a human hair follicle illustrating the typical cytology of a retaining cell, i e a cell synthesizing protein but not secreting it, to be contrasted with the cells of Fig 4 (shown at smaller magnification) The same dense RNA containing particles are present but these are 'free' and not associated with membranes Where memb...

FIG 6 A ... xation
region of a r... Golgi
Birbeck ... S C

FIG. 11 Cross-section of the plasma membrane of *Amoeba proteus* (See Fig 13)

PLATE III

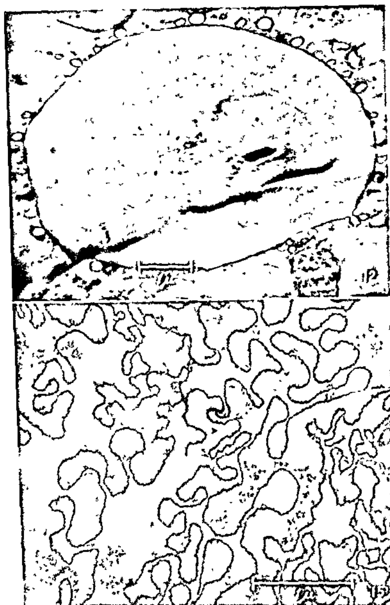


Fig 12 Section of a food vacuole in the absorption stage in an amoeba Undigested residues and small vacuoles budding-off the larger (Fig 9c) are to be seen Fixed in osmium tetroxide

Fig 15 Section of a crumpled bundle of a layer of egg albumen spread on water (see text for explanation) Fixed in permanganate

PLATE IV

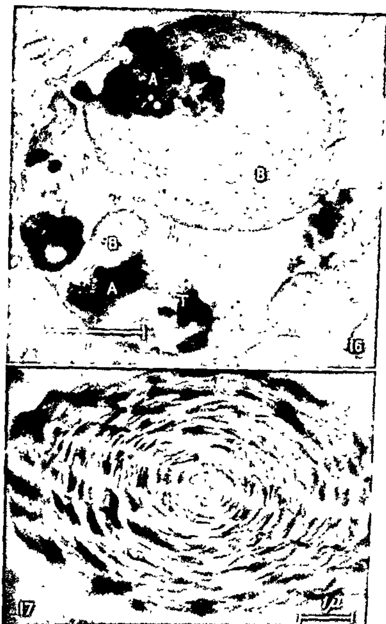


FIG. 16 A phospholipid containing vacuole in a developing germinal cell from the ovotestis of *Helix pomatia*. Notice at A amorphous material, B masses of parallel membranes (myelinic form), at T tubules presumably formed from the same phospholipid as B, but representing an alternative structural aspect.

FIG. 17 Cross-section of an artificially formed myelinic figure to show concentric disposition of the phospholipid sheets. *Cf.* B in Fig. 16. At high resolution the condensed areas show a periodicity of 40Å. Osmium fixation.

Membrane dynamics

Current speculation is much concerned with the idea that a good deal of cellular activity will be explained in terms of the pleomorphism of membrane systems, that is, their ability to transform readily from one form to another. The following simple 'basic operations' seem capable of extensive application:

(*a*) sac ⇌ vesicle transformation (Fig. 9a)
(*b*) evaginations and invaginations as tubules or pleats (Fig. 9b)
(*c*) budding off or absorption of vesicles (Fig. 9c)

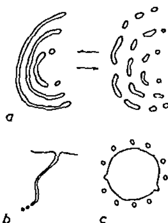


FIG. 9 Membrane dynamics based on pleomorphic changes: (*a*) the transformation vesicles—flattened sacs which occur when microsomes (Plate I, Fig. 3) are produced from the basophilic reticulum and which probably also occur *in vivo*; (*b*) involution of the plasma membrane as a pleat consisting of two closely opposed surfaces with the separation of vesicles from the end (β-cytomembranes), (*c*) budding off of small vesicles from larger (Plate II, Fig 11) or, conversely, the coalescence of small vesicles to form larger (contractile vacuole).

These operations may be used to transfer materials across membranes or to transport them about the cell. Their molecular mechanism is obscure but might well have parallels in the phenomenon of emulsification. That is, the change in surface curvature and area may be brought about by the absorption or desorption of surface-active substances, e.g. proteins (Fig. 10).

By specifically absorbing certain substances on special sites on membranes, these substances may be selectively transferred, adsorbed or transported.

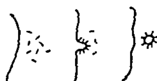


FIG. 10. Hypothesis to explain some of the phenomena shown in Fig. 9, based on the reversible absorption of surface-active molecules on membranes leading to their emulsification.

In the study of cellular dynamics of this kind, electron microscopy is definitely at a loss, since it deals with a fixed state and provides only a snapshot of that state. There can be no question of following a process in a particular cell as it develops in time. The difficulty can be got around by examining a succession of states in different cells; that is, by means of light microscopy or in some other way, the times of fixation are selected to record a succession of events. For example, the course of food absorption and digestion in a large cell, such as an amoeba, may be easily controlled by light microscopy and stages corresponding to capture, vacuolization, digestion, and absorption may be chosen. An electron micrograph of the absorption stage is shown in Plate III, Figure 12. The reduction of the large vacuole by the budding-off of small vesicles is evident.

The pleomorphism of membranes is certainly associated with their molecular composition and fine structure, which can be studied to a limited degree electron microscopically. High-resolution micrographs reveal a lamella structure consisting of two dense layers enclosing a lighter (Plate II, Fig. 11). Often further material adsorbed external to the denser layers may be noted. Physicochemical studies on the conductivity of membranes and their permeability suggest a predominately lipid character (Davson and Danielli, 1943). Chemical analysis reveals both lipid (phospholipids) and protein. The hypothesis of the lipoprotein sandwich, largely developed by Danielli (1942), combines these findings and is in accord with the electron microscopic

evidence if we assume that the fixatives (both osmium and permanganate) reacted mostly with the outer layers of the sandwich (protein plus the active groups of the lipid), see



FIG. 13. The Danielli model of the plasma membrane. P = protein; L = lipid layer (cf. Plate II, Fig 11).

Figures 13 and 14. We are, however, far from possessing a precise enough picture of this structure to predict with confidence its dynamic behaviour.

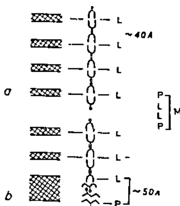


FIG. 14. Artificial myelinic forms and their molecular interpretation. (*a*) The alternation of light and dark bands (periodicity ~ 40 A) noted in a cross-section of an osmium fixed myelinic tubule. On the right-hand side is shown an interpretation in molecular terms in which the dark bands are located in the lipid regions. Alternatively the osmium deposit may be thought to occur at the hydrophilic ends of the molecules (⫯) giving the same periodicity (*b*) The surface of a myelinic tubule formed ...

Model structures formed from membranes

The use of artificial models of biological structures to examine the action of fixatives, etc., was mentioned earlier. The approach

promises to be particularly valuable in the investigation of the difficult problems associated with cytomembranes. The two classes of substances likely to be implicated in the construction of membranes are proteins and phospholipids and a great deal of information concerning their structural possibilities, from sources other than electron microscopy, already exists (Goldacre, 1958; Nageotte, 1936). Recent papers by Revell, Ito and Fawcett (1958) in the United States, Stoeckenius in Germany (1958) and myself (1957 and 1958) show a growing interest by microscopists.

Protein films of the order of thickness of biological membranes can be made most easily by spreading at an air-water interface in a Langmuir trough. The spread film may be swept up to one end of the trough and collected as a thin thread ready to be processed for electron microscopy by fixing, staining and embedding. Sections cut at right angles to the length of the thread reveal cross-sections of the protein membrane (Plate III, Fig. 15). Such experiments show that osmium tetroxide and potassium permanganate stain strongly, formaldehyde and alchol-fixed films, at first barely visible, stain strongly with phosphotungstic acid and uranyl acetate, both commonly used stains. The behaviour of phospholipids is of greater interest since the earlier work of light microscopists and crystallographers (Schmitt, Bear and Palmer, 1941; Finean, 1953) has established a close identity between the myelinic forms produced spontaneously *in vitro* by these materials and the membranous sheaths of nerve fibres.

Phospholipids can be extracted from brain (Nageotte, 1936) in quantities and have been used mostly in these experiments. The cell structures from which these lipids are removed are probably the myelin sheaths of the processes of the neurons. When placed in water, a small lump of the extract immediately begins to sprout birefringent myelinic forms (Plate IV, Fig. 17) whose variety and behaviour have interested many observers. Observations by means of polarized light and X-ray diffraction showed that the long tubules consisted of concentric shells of bimolecular leaflets of the phospholipid molecules (Fig. 14). Since brain phospholipids react strongly with osmium tetroxide, these myelinic forms may be fixed, embedded and examined

electron microscopically. Cross-sections reveal, as expected, concentric circles of alternating light and dense layers. Undoubtedly these represent sections of successive layers of phospholipid, but it is by no means immediately obvious what part of the phospholipid complex is actually stained by the osmium to form the surface of enhanced density.

The most accurate measurements of the dimensions of these structures are those of Stoeckenius, whose remarkable micrographs were displayed at the recent Conference on Electron Microscopy at Berlin (September 1958). In Figure 14 the dimensions (as found) are shown and on the right-hand side is his interpretation of the image in molecular terms.

When the myelinic forms are developed in a protein solution, the superficial layer is seen to be darker and thicker (Fig. 14b) and this would seem to indicate the adsorption of protein. Again the exact interpretation is not clear, but it seems likely that this outside layer is analogous to *half* a single free membrane as explained at M in Figure 14, in which the symbols make a minimum of assumptions as to exact structure.

In these preparations one may find by chance formations resembling cell organelles. Although the conditions governing their appearance are not known, this at least demonstrates something of the structural potentialities of phospholipids and proteins.

Another way of studying these potentialities is to take advantage of the fact that some cells appear to store lumps of the material in special vacuoles (Chou, 1957). In the germinal cells of the ovotestis of *Helix* one may observe numbers of such vacuoles in which the phospholipid complex (?) can be seen giving rise to a remarkable series of parallel membrane structures, tubules and vacuoles (Plate IV, Fig. 16). Looking at such forms one can, in at least a geometrical sense, conceive that the material offers the means of forming a great variety of structural elements fully adequate to account for the membranous organelles actually observed. We can observe in particular the spontaneous formation of the two essential 'bricks' of membranous organelles: the double membrane and the tubule (Plate IV, Fig. 16, at B and T).

promises to be particularly valuable in the investigation of the difficult problems associated with cytomembranes. The two classes of substances likely to be implicated in the construction of membranes are proteins and phospholipids and a great deal of information concerning their structural possibilities, from sources other than electron microscopy, already exists (Goldacre, 1958; Nageotte, 1936). Recent papers by Revell, Ito and Fawcett (1958) in the United States, Stoeckenius in Germany (1958) and myself (1957 and 1958) show a growing interest by microscopists.

Protein films of the order of thickness of biological membranes can be made most easily by spreading at an air-water interface in a Langmuir trough. The spread film may be swept up to one end of the trough and collected as a thin thread ready to be processed for electron microscopy by fixing, staining and embedding. Sections cut at right angles to the length of the thread reveal cross-sections of the protein membrane (Plate III, Fig. 15). Such experiments show that osmium tetroxide and potassium permanganate stain strongly, formaldehyde and alchol-fixed films, at first barely visible, stain strongly with phosphotungstic acid and uranyl acetate, both commonly used stains. The behaviour of phospholipids is of greater interest since the earlier work of light microscopists and crystallographers (Schmitt, Bear and Palmer, 1941; Finean, 1953) has established a close identity between the myelinic forms produced spontaneously *in vitro* by these materials and the membranous sheaths of nerve fibres.

Phospholipids can be extracted from brain (Nageotte, 1936) in quantities and have been used mostly in these experiments. The cell structures from which these lipids are removed are probably the myelin sheaths of the processes of the neurons. When placed in water, a small lump of the extract immediately begins to sprout birefringent myelinic forms (Plate IV, Fig. 17) whose variety and behaviour have interested many observers. Observations by means of polarized light and X-ray diffraction showed that the long tubules consisted of concentric shells of bimolecular leaflets of the phospholipid molecules (Fig. 14). Since brain phospholipids react strongly with osmium tetroxide, these myelinic forms may be fixed, embedded and examined

electron microscopically. Cross-sections reveal, as expected, concentric circles of alternating light and dense layers. Undoubtedly these represent sections of successive layers of phospholipid, but it is by no means immediately obvious what part of the phospholipid complex is actually stained by the osmium to form the surface of enhanced density.

The most accurate measurements of the dimensions of these structures are those of Stoeckenius, whose remarkable micrographs were displayed at the recent Conference on Electron Microscopy at Berlin (September 1958). In Figure 14 the dimensions (as found) are shown and on the right-hand side is his interpretation of the image in molecular terms.

When the myelinic forms are developed in a protein solution, the superficial layer is seen to be darker and thicker (Fig. 14b) and this would seem to indicate the adsorption of protein. Again the exact interpretation is not clear, but it seems likely that this outside layer is analogous to *half* a single free membrane as explained at M in Figure 14, in which the symbols make a minimum of assumptions as to exact structure.

In these preparations one may find by chance formations resembling cell organelles. Although the conditions governing their appearance are not known, this at least demonstrates something of the structural potentialities of phospholipids and proteins.

Another way of studying these potentialities is to take advantage of the fact that some cells appear to store lumps of the material in special vacuoles (Chou, 1957). In the germinal cells of the ovotestis of *Helix* one may observe numbers of such vacuoles in which the phospholipid complex (?) can be seen giving rise to a remarkable series of parallel membrane structures, tubules and vacuoles (Plate IV, Fig. 16). Looking at such forms one can, in at least a geometrical sense, conceive that the material offers the means of forming a great variety of structural elements fully adequate to account for the membranous organelles actually observed. We can observe in particular the spontaneous formation of the two essential 'bricks' of membranous organelles: the double membrane and the tubule (Plate IV, Fig. 16, at B and T).

A more complete study, preferably *in vitro*, of the behaviour of cellular constituents will certainly increase confidence in our ability to deduce the nature of original structures from electron micrographs, nevertheless there will remain much morphological detail which can neither be confirmed by an alternative experimental technique nor, in its particulars, be inferred from a general knowledge of the chemical behaviour of its components. To return to the point of view of the introduction, the validity of such images is to be assessed by the contribution they make towards producing an integrated picture of the cell. They bring 'genuine' information, not because any other experimental approach confirms it, but because it helps to make sense when combined with all other sources of information. Information theory teaches us that no data have any other sort of validity but that it is entirely adequate for the construction of scientific theories.

ACKNOWLEDGEMENTS

I am gratef [illegible]
lems with [illegible]
M. S. C. B [illegible]
to Mr. M. J. Docherty who made the photographic enlargements.

REFERENCES

BAHR, G. F. (1954) *Exp. Cell Res.* 7, 457.

BAKER, J. R. (1945). *Cytological Technique* Methuen, London.

BIRBECK, M. S. C. and MERCER, E. H (1957a) *Brit. J. appl. Phys* 8, 259.

BIRBECK, M. S. C and MERCER, E H. (1957b) *J. R. micr. Soc* 76, 159.

BIRBECK, M. S. C. and MERCER, E. H. (1957c). *J. biophys biochem Cytol.* 3, 203.

BRILLOUIN, L. (1956). *Science and Information Theory* Academic Press Inc., New York.

[illegible]

[illegible] e. Oxford University Press, London.

DAVSON, H. and DANIELLI, J. F. (1943). *The Permeability of Natural Membranes* Cambridge University Press.

FINEAN, J. B. (1953) *Exp Cell Res.* 5, 202.

GLAUERT, A. M., ROGERS, G. E. and GLAUERT, R. M. (1956). *Nature, Lond*, 178, 803.
GOLDACRE, R. J. (1958). In *Surface Phenomena in Chemistry and Biology*, Pergamon Press, London, p. 278.
HAGENAU, F. (1958). *Int. Rev. Cytol.* 7, 425.
KELLENBERGER, E., SCHWAB, W. and RYTER, A. (1956). *Experientia*, 12, 421.
[illegible]

In press.
NAGEOTTE, J. (1936). *Actualités Scientifiques et Industrielles*. Hermann, Paris.
PALADE, G. E. (1952a). *J. exp. Med.* 95, 285.
PALADE, G. E. (1952b). *Anat. Rec.* 114, 427.
PALADE, G. E. (1955). *J. biophys biochem. Cytol.* 1, 567.
PALADE, G. E. and SIEKEVITZ, P. (1956). *J. biophys. biochem Cytol.* 2, 171.
PETERS, R. A. (1937). In *Perspectives in Biochemistry*, eds. J. Needham and D. E. Green, Cambridge University Press, p. 36.
PORTER, K. R. (1957). The Harvey Lectures (1955–6), Series 51, 175 Academic Press Inc, New York.
REVELL, J. P., ITO, S. and FAWCETT, D. W. (1958), *J. biophys. biochem. Cytol.* 4, 495.
ROBERTSON, J. R. (1957). *J biophys. biochem. Cytol.* 3, 1043.
RUSSELL, BERTRAND (1958). *The Listener*, 59, 223.
SCHMIDT, W. J. (1937). *Die Doppelbrechnung von Karyoplasma, Zytoplasma und Metaplasma*. Gebruder Borntraeger, Berlin.
SCHMITT, F. O., BEAR, R. S. and PALMER, K. J. (1941). *J. cell. comp. Physiol* 18, 31.
SJÖSTRAND, F. S. (1953). *Nature, Lond.* 171, 30.
STOECKENIUS, W. (1960). Int. Confer. on Electron Microscopy, Berlin, 1958. In press
STRANGEWAYS, T. S. P. and CANTI, R. G. (1927). *Quart. J. micro. Sci.* 71, 1.

IV

Intersexuality

C. N. ARMSTRONG

I SHOULD define intersex as that *condition in which, in the* same individual, there are male and female characteristics to an abnormal degree. In former years, the recognizable sexual characteristics in man were external body form, external genitalia and hair distribution; later, the histology and sex of the gonads was added; today, nuclear sex, and tomorrow, perhaps, sexual behaviour, the psychological sex.

I *would say the criteria of sex are four:*

1. chromosomal sex, i.e. the nuclear sex;
2. *sex of the gonads, i.e. ovaries or testes;*
3. body sex, external genitalia and form. (What Professor Jost (1958) of Paris describes as the Nursery sex which becomes the Legal sex. In Britain, it would be more correct to say the Registered sex—there is no legal definition of sex in this country);
4. psychological sex or sexual behaviour.

I do not include hormone secretion, response of the target organs and environmental factors because they are more correctly the mechanisms *by which the sexual characteristics are* developed or influenced rather than the final sexual features or criteria.

I include sexual behaviour, which is psychological sex, in the criteria of sex because I am of the opinion that sexual behaviour is instinctive on a genetic and constitutional basis, and not wholly dependent upon environmental influences, and, therefore, *I include in intersexuality those conditions in* which the

psychological sex and the recognizable criteria of physical sex in an individual are not the same sex.

The two important psychological intersex syndromes are:

(1) transvestism, which is quite distinct from homosexuality, and in which there are likely to be aetiological constitutional factors, probably on a genetic basis. I dealt with this condition in my contribution to the 'Symposium on Nuclear Sex' held in September 1957, and time does not allow any further reference in this lecture (Armstrong, 1958);

(2) homosexuality, which is the other abnormality of sexual behaviour. In these individuals libido is directed towards a person of the same physical sex, and should be recognized as distinct from merely the performance of abnormal sexual acts. There is direct evidence from studies of homosexuality, and indirect evidence from psychological studies, of a genetic factor in the differentiation of psychological sex orientation.

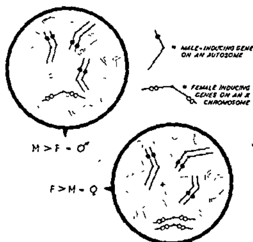
There is need for research in constitutional psychological sex abnormalities into the possible physical aetiological factors: genetic, constitutional, and hormonal, and I do not intend to refer again to transvestism and homosexuality after having explained why I include them in Intersex.

Grumbach and Murray Barr (1958) state their opinion that all those individuals in whom the chromosomal sex is contrary to the phenotypic sex (physiological sex) should be regarded as intersexes.

Normal Sex Determination, Differentiation and Development

The various problems of intersex can only be approached by clearly appreciating normal sex determination, differentiation, and development. The sex of an individual is determined by chromosomal factors at the time of fertilization when the spermatozoon fertilizes the ovum. This is the chromosomal sex. If subsequent development is normal, the gonads, the external genitalia, the body form and sexual behaviour are masculine or feminine in accordance with the chromosomal or genetic sex. Thus the genetic sex and phenotypic sex (physiological sex) are the same sex.

The factors in development are chromosomal, genic and endocrine. In man each cell has 46 chromosomes, 2 sex chromosomes (two Xs or an X and a Y) and 22 pairs of autosomes. There are female-inducing genes in the X female chromosomes and male-inducing genes in the autosomes in the proportion normally that the male-inducing genes in the autosomes are greater than the female-inducing genes in one X chromosome, but less than in two X chromosomes. All ova are X. Thus an ovum fertilized by an X spermatozoon develops a female balance, and an ovum fertilized by a Y spermatozoon develops a male balance. For convenience, an arbitrary method of representing this is to regard an X chromosome as 1 F and autosomes as $1\frac{1}{2}$ M: thus an XX ovum = 2 F : $1\frac{1}{2}$ M, female factors predominating, and an XY ovum = 1 F : $1\frac{1}{2}$ M, male factors predominating, or X = 4, autosomes 6, thus XX = 8 F : 6 M, and XY = 4 F : 6 M. Y plays no part so far as is known. (Fig. 1.)



[illegible] that the female-inducing genes are carried in [illegible]

Chromosomal and nuclear sex

Except under unusual circumstances, it was not possible to detect the chromosomal sex of a person until in 1949 Murray

Plate V

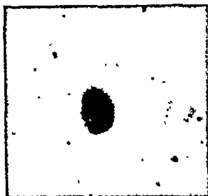


Fig 2 Cell from oral mucosa (× 1200) showing special mass of chromatin (chromocentre) adjacent to the nuclear membrane



Fig 4 *Neutrophil leucocyte (× 1800) showing the projection or 'club' found in females*

Barr discovered that there was a special mass of chromatin (chromocentre) which is characteristic of female cells. This was a discovery of very great importance in the study of intersexuality, and has been the means of clarifying many of the problems. Murray Barr observed in nervous tissue an intranuclear satelli[illegible]
other [illegible]
Fig. 2, [illegible]
80 per cent of nuclei in females as a small body adjacent to the nuclear membrane compared with only 5 per cent of a similar structure in males. The question arose whether this was a secondary sex characteristic through hormone or other influences, but conclusive evidence suggests that the presence or absence of sex chromatin is not a secondary sex characteristic, but a basic property of somatic cells according to the individual's original sex in the early embryo, regardless of any deviation that [illegible]

[illegible]

chromatic regions of the two X chromosomes. Lennox and his colleagues (1958) have shown recently that the two X chromosomes extend from the nucleolus to the membrane of the nucleus and the heterochromatic region is adjacent to the membrane (Fig. 3).

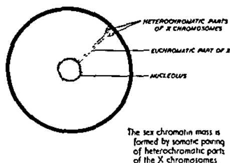


FIG. 3. Drawing showing how the two X chromosomes extend from the nucleolus to the membrane of the nucleus with the heterochromatic region adjacent to the membrane.

In 1954, Davidson and Robertson Smith (1955) found a projection or 'club' in normal average 6 out of 200 neutrophil leucocytic nuclei from females and a normal minimum 6 in 500 and none in those from males (Plate V, Fig. 4). These clubs may be a secondary sex characteristic and the percentage appears to be less than normal in certain cases of intersex who are genetic females, e.g. I have found this to be so in true hermaphroditism of XX constitution and in a female homosexual who attended my outpatients, an extremely rare and unusual occurrence. Davidson and Robertson Smith report it in Klinefelter's syndrome of XX constitution.

Sex determination by chromosomes, which is a genetic phenomenon, has to be distinguished from sex differentiation which is a hormonal process dependent mainly upon the development or non-development of a testis. Before seven weeks, the gonad of the foetus is undifferentiated and acquires two major somatic components, the cortex and the medulla, in addition to the germ cells. The cortex plays the major role in ovarian differentiation, the medulla in testicular differentiation.

In the ovary, the cortex increases in size and eventually becomes a typical ovary. In the testis the cortex degenerates, and in the medulla the seminiferous tubules develop and there are changes of special interest in the interstitial cells. According to Jost (1958) genetic sex factors are responsible for the prevalence of either the cortex or the medulla. The genital tract differentiates later than the gonads. The genital systems develop from two ducts, the mesonephric urinary duct, the Wolffian duct, and another duct which develops on the anterior and external margin of the mesonephros, the Mullerian ducts. In the female, the Wolffian ducts do not become sexual structures, they retrogress, while the Mullerian ducts persist and differentiate into tubes and uterus In the male, the Mullerian ducts retrogress while the Wolffian ducts become part of the sexual system with development of vas deferens and epididymis. The medulla of the gonad gives rise to seminiferous tubules. The testicular interstitial cells of the male foetus become hypertrophied and show signs of activity. Jost has shown by animal experiment that the endocrine secretion of the foetal testis governs further

... he Foetal Testis is ... d to apply this to ... it simplifies the understanding of the development of the Intersex syndromes.

NORMAL

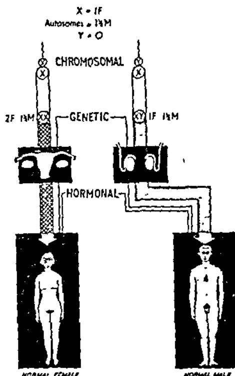


FIG. 5 Diagram illustrating the chromosomal, genetic, and hormonal factors in normal sexual development, and that it is the foetal testis which governs masculine development, which is a deviation from the normal basic female

Female sex the basic or neutral sex

Castration of the female foetus does not interfere with feminine differentiation of the female rabbit foetus; ovarectomized females develop a complete feminine genital tract. Thus feminine

organogenesis does not depend upon the presence of the ovaries. Castration of male foetus shows the primordial importance of the testes as body-sex differentiators. In foetus castrated before initiation of somatic sexual differentiation, no male characteristics develop and the whole genital tract becomes feminine similar to that observed in castrated female foetus. The Wolffian ducts disappear, the Müllerian ducts are retained and differentiate into tubes, uterine horns and Müllerian vagina: the urogenital sinus and the external genitalia become feminine. Jost has therefore shown by his experimental work that the 'neutral' or 'gonadless' type developed in the absence of any sex gland (or in the presence of ovaries) is feminine. It is the foetal testis which prevents male embryo from acquiring a female body by suppressing very early the Mullerian ducts and by stimulating masculinization of the Wolffian ducts and the resultant development of masculine structures. This is of very great importance: it proves that the basic or neutral sex is female and the development of a male body is a deviation from the normal basic course of development. If this fact is appreciated, it becomes easier to understand how the intersexes develop.

The Intersexes due to testicular deficiency

In other experiments aimed at partial foetal testicular deficiency, it has been possible to influence the genital tract to a reduced extent. *Castration, if carried out early in foetal life, results in feminine external genitalia, if carried out later,* various underdevelopment of masculine characteristics results according to the time and stage at which it was carried out, the last being hypospadias only; and finally, castration, if late, does not influence the development of normal male characteristics. Also it has been shown that unilateral castration results in a lateral asymmetry of the genital tract, one side being normally masculine and the other, the castrated side, feminine. This indicates that the activity of the foetal testis is to some extent spatially restricted.

[illegible]

of [illegible]

or [illegible]

have already said, are chromosomal, genic and endocrine. I intend to suggest at what level these factors are effective in the development of the main classical intersex syndromes:

INTERSEX (XX)

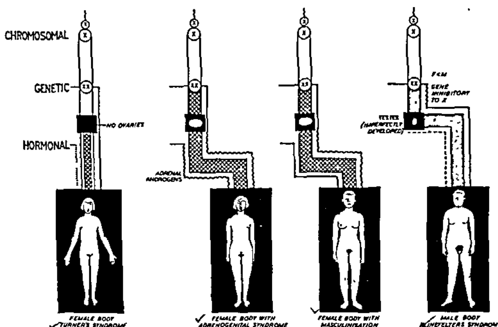


FIG 6. Diagram illustrating the genetic and hormonal factors in the development of intersex in individuals of XX constitution, genetic females, and that the basic or 'gonadless' sex is female.

EXAMPLES OF INTERSEX IN GENETIC FEMALES

Chromosomal or genetic females—XX *constitution* (Fig. 6)

(1) *Turner's Syndrome* (gonadal dysgenesis: ovarian agenesis). In this syndrome described by Turner (1938), in its typical form, the body is female and often associated with multiple congenital anomalies (e.g. webbed neck), external female characteristics, some pubic and axillary hair, sexually infantile. Gonadotrophins high. No gonad is present, only a narrow ridge

in the region of the ovary. *The nuclear sex may be female or male:* male in the majority of cases.

[illegible]

tract with development of a penis-like phallus, masculine hirsutism, absent breast development, and primary amenorrhoea. It is probably genetically determined producing an inborn error of metabolism in which there is a defect in the *production of hydrocortisone by the adrenals; the deficiency* of hydrocortisone leads to an increased secretion of corticotrophin by the pituitary which produces hyperplasia of the adrenals, and an increased output of androgens. (Diagnosed by determining nuclear sex and urinary 17-ketosteroid excretion and treated successfully, as Lawson Wilkins (1950) has shown, by cortisone administration.)

(3) *Masculinization can also obviously occur from excess of androgen from other sources, i.e. without evidence of abnormal secretion in the patient*, e.g. the mother, or if the mother during pregnancy has been given progesterone, which is sometimes administered for the treatment of threatened abortion (Lawson Wilkins, 1958). As the androgen effect comes in late, the gonads in all these cases are ovaries and the abnormalities of other sexual organs and body form will depend upon the amount of androgen and the time when introduced.

(4) *Masculinization without evidence of abnormal endocrine secretion or androgens from maternal sources.* Rare, probably of genetic determination (e.g. more masculine influence on autosomes). A case has been reported by Armstrong (1955a).

(5) *Klinefelter's syndrome*[1] (primary micro-orchidism, seminiferous tubule dysgenesis). This syndrome originally described by Klinefelter (1942) includes female nuclear sex, gynaecomastia

[1] Since this lecture was delivered, Ford and his colleagues have shown it is prob- [illegible] constitution in Klinefelter's syndrome is X X Y with a [illegible] total of 45 chromo- [illegible] xtra Y, or a male [illegible] has lost a Y or a [illegible] rries some masculinizing genes.

and atrophic testes, atrophy and hyalinization of the seminiferous tubules, Leydig cells are abundant, and there is usually azoospermia, but in some cases spermatogenesis has been observed and may be fertile (Ferguson-Smith and Munro, 1958). The aetiology has not yet been determined. It has been reported in identical twins (Grumbach and Barr, 1958) and is probably at genetic level. Possible explanations are:

(i) Lebedeff in 1944 described a gene 'ix' in drosophila which produces male morphology in XX insects. Thus one can speculate that in humans such a gene in an XX embryo would inhibit the female-inducing factors and give the predominance to the male-inducing factors in the autosomes. This would result in male balance causing the development of a testis and subsequent male anatomy, but as such a testis is not likely to be fully normal, might result in some of the feminizing characteristics, and absence of Y may be responsible for deficient spermatogenesis;

(ii) the syndrome may result from a compensatory development of the gonadal medulla of the undifferentiated gonad following a defect of the cortex;

[illegible]

c

c

s

composition would bring about the development of a normal testis. A normal number of XX sex cells (i.e. twice the quantity of X in the normal male) would result in a normal ovary. It could be presumed from this that an intermediate number of XX sex cells would give rise to a fairly normal testis which, I suggest, would be sterile if Y contains a gene necessary for maturation of the testes. Following the same arguments, the relative deficiency of XX cases of gonadal agenesis (Turner's syndrome) is balanced by the XX cases of Klinefelter's syndrome.

The case against Witschi's attractive theory is that if the sex cells in the frog are destroyed by irradiation, the animal still develops an ovary or a testis—as demonstrated by Burns (1955). The important fact to appreciate is that development of the

testes causes the formation of a male somatic body in a chromosomal female.

Examples of Intersex in Genetic Males

XY Constitution (Fig. 7)

INTERSEX (XY)

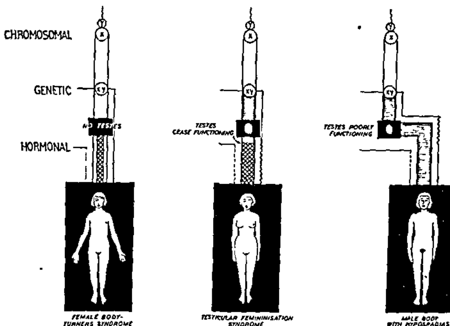


Fig. 7. Diagram illustrating the genetic and hormonal factors in the development of intersex in individuals of XY constitution, genetic males, that the basic sex is female, and the degree of masculine development depends upon the functioning of the foetal testis If it ceases to function very early, a female body develops, and, if later, or poorly functioning, female micturition will be the only feminine characteristic

(1) *Turner's Syndrome* (gonadal dysgenesis). In this syndrome the typical form is as already described in genetic females: external female characteristics, sexually infantile; multiple congenital anomalies are frequently present such as webbed neck, wide carrying angle, etc. No gonad is present. The importance

of this syndrome lies in the fact that a genetic male who does not form a testis develops a female body—illustrating that the female is the basic sex.

(2) *Testicular Feminization Syndrome.* This syndrome, cases of which have been reported by Goldberg and Maxwell (1948); Schneider, Omman and Hoerr (1952); Morris (1953); Wilkins [illegible]

ends in a blind pouch, there is no uterus; the nuclear sex is male and there are intra-abdominal testes. The level of oestrogen and androgen excretion is about normal and it is reported that castration results in a fall in urinary oestrogens. The aetiology is no doubt genetic, and various genetic mechanisms have been suggested: sex-linked recessive, or a sex-limited dominant gene, or an aberration of the sex chromosomes in which affected individuals are postulated to have an XXY sex chromosome constitution. An important feature of these patients is that administration of androgen does not cause growth of pubic hair (Lawson Wilkins, 1957) which raises the whole question of response of the target organs to hormones and whether, during development, a part of the body, or section of an organ, becomes sensitive to a specified hormone while other parts have no sensitivity. The sensitivity of tissues to hormones and possible local anti-hormone activity requires further study.

It is obvious that, dependent upon the stage of intra-uterine life at which the testis ceases to function normally, variants will result ending in:

(3) *Male with hypospadias*, in which the only feminizing feature is the female micturition. In my experience, such hypospadias patients brought up as girls have male psychological sex and experience no psychological difficulties in changing their social sex to male.

True Hermaphroditism

In these cases, one gonad is an ovary and the other a testis or ovotestis, or both gonads are ovotestes. The external genitalia may vary towards male or female characteristics. There must be a cortico-medullary imbalance during the differentiation of

testes causes the formation of a male somatic body in a chromosomal female.

Examples of Intersex in Genetic Males

XY Constitution (Fig. 7)

INTERSEX (XY)

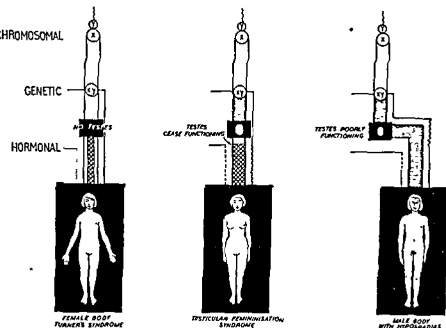


FIG 7. Diagram illustrating the genetic and hormonal factors in the development of intersex in individuals of XY constitution, genetic males, that the basic sex is female, and the degree of masculine development depends upon the functioning of the foetal testis. If it ceases to function very early, a female body develops, and, if later, or poorly functioning, female micturition will be the only feminine characteristic.

(1) *Turner's Syndrome* (gonadal dysgenesis). In this syndrome the typical form is as already described in genetic females: external female characteristics, sexually infantile; multiple congenital anomalies are frequently present such as webbed neck, wide carrying angle, etc. No gonad is present. The importance

of development of some of the classical syndromes, and the possibility of variants. The recent discovery by Murray Barr of an easy method of determining the chromosomal sex, and recent advances in modern medicine, open a broad and exciting field for further research in the fascinating unsolved problems of Sex and Intersex.

HYPOTHETICAL GENETIC CAUSE OF TRUE HERMAPHRODITISM

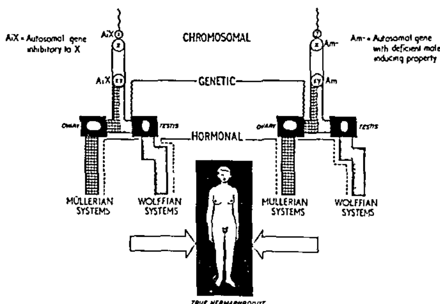


FIG. 8. Diagram illustrating the possible genetic and hormonal factors in the development of true Hermaphroditism.

SUMMARY

(1) Intersex is defined as that condition in which, in the same individual, there are male and female characteristics to an abnormal degree.

(2) The criteria of sex are regarded as being (i) chromosomal sex; (ii) sex of the gonads; (iii) body sex; external genitalia and form; (iv) psychological sex. Reasons are given for including

either one or both gonads. The aetiology is unknown. Possible explanations are:

(i) The unilateral influence of the testis may be an explanation *of some cases of true hermaphroditism if there was failure* of one testis only, and feminine development on that side only, and it must be recognized that the development of the body, in any case, is not always symmetrical.

(ii) Witschi's suggestion that the primordial sex cells act quantitatively, so that spatially irregular migration of the primordial sex cells (which have an extra-gonadal origin and migrate to the gonadal site), too many to one side, too few to the other, or too many to one pole and too few to the other in both sides, might result in the development of a testis or an ovary on one or other side, depending upon the number of the sex cells on either side and ovotestes depending upon polar distribution.

(iii) In any case, the effect of X is no doubt quantitative whether a small dose of X stimulates the medulla and a large dose stimulates the cortex, or a small dose of X permits the early differentiation of the gonad and a large dose of X delays gonadal differentiation; always remembering that an early differentiation must be a testis, a late differentiation an ovary.

Whichever is true, the quantitative action of X allows a multitude of genetic mechanisms whereby the value of X, or XX, could fall nearly equally between normal X and normal XX, to allow a mixture nearly equally between normal X and normal XX and thus allow a mixture of testicle and ovary on the two sides.

The value of X can be diminished by the presence of an anti-X autosomal gene A_x, or by deletion of part of X, or the value of X can be increased by reduplication of a segment of X or by the presence of an autosomal gene with deficient male-inducing property A_{m-} (Fig. 8).

The result in either case could be Hermaphroditus.

I have endeavoured to throw some light on the whole spectrum of Intersex by focusing attention on the chromosomal, genic and hormonal factors concerned in the development of intersex with the object of attempting to explain the mechanism

BARR, M. L. (1957). *Brit. J. Urol.* 29, 3.
BARR, M. L. (1957). *Progress in Gynaecology*, 3, 131.
BARR, M. L. and BERTRAM, E. G. (1949). *Nature, Lond.* 163, 676.
[illegible]
DAVIDSON, W. M. and SMITH, D. R. (1955). *Postgrad. med. J.* 32, 578.
DAVIDSON, W. M. and SMITH, D. R. (1958). *Symposium on Nuclear Sex.* Interscience Pub. Inc. p. 93.
FERGUSON-SMITH, M. A. (1958). *Lancet*, p 928.
FERGUSON-SMITH, M. A., LENNOX, B., MACK, W. S. and STEWART, J. S. S. (1957). *Lancet*, p 167.
FERGUSON-SMITH, M. A. and MUNRO, I. B (1958) *Scot. med. J.* 3, 39.
FORD, C E., JONES, K. W., MILLER, O. J., MITTWOCH, U., PENROSE, L. S., RIDLER, M and SHAPIRO, A. (1959) *Lancet*, I, 709.
FORD, C. E., JONES, K W., POLANI, P. E., DE ALMEIDA, J. C. and BRIGGS, J. H. (1959) *Lancet*, I, 711.
GOLDBERG, M. B. and MAXWELL, A. F. (1948). *J. clin. Endocr* 8, 367.
GRUMBACH, M. M. and BARR, M. L. (1958). *Recent Progr. Hormone Res.* 14, 255
JOST, A (1958) *Embryonic Sexual Differentiation.* Williams & Wilkins Co. Baltimore Chap 2
KLINEFELTER, H F., REIFENSTEIN, E. C. and ALBRIGHT, R. (1942). *J. clin. Endocr.* 2, 615.
LEBEDEFF, G. A (1934) *Proc. Nat Acad. Sci* 20, 613.
LENNOX, B. (1956). *Scot med J* 1, 97.
LENNOX, B , FERGUSON-SMITH, M A., MACK, W. S. and STEWART, J S. S. [illegible]
[illegible]
Endocr 12, 423.
STEWART, J. S. S., FERGUSON-SMITH, M A , LENNOX, B. and MACK, W S. (1958) *Lancet*, 2, 117
SWYER, G I. M (1955) *Brit med. J* II, 709.
THOMPSON, B K , HAGGAR, R. A and BARR, M. L. (1957). *J comp. Neurol* 108, 2
TURNER, H H (1938) *Endocrinology*, 23, 566.
WILKINS, L (1950) *The Diagnosis and Treatment of Endocrine Disorders in Childhood and Adolescence* C. C. Thomas, Springfield, Illinois.
WILKINS, L. (1957). *The Diagnosis and Treatment of Endocrine Disorders in Childhood and Adolescence* 2nd edn. C. C. Thomas, Springfield, Illinois.
WILKINS, L (1958) Personal communication

psychological sex in the criteria of sex and for including transvestism and homosexuality in intersex.

(3) Normal *sex determination, differentiation and development* are as described. Attention is drawn to the fact that the nuclear chromocentre reveals a female chromosome constitution. The influence of genetic factors and hormones on sexual differentiation and development are discussed.

(4) Professor Jost's experimental work has proved that the basic or neutral sex is female and the foetal testis is the body sex differentiator.

(5) I have regarded the development of the male body as a deviation from the normal basic course of development and applied this to the study of the classical intersexes, all of which are reviewed, and explanations as to how some of the intersexes are due to complete or partial foetal testicular failure are discussed.

(6) Theories of abnormal genetic factors concerned in the development of certain intersex syndromes are mentioned and suggested.

(7) Attention is drawn to the fact that many of the problems of intersex, physical and psychological, are as yet unsolved and there is a broad field for further research.

ACKNOWLEDGEMENTS

[illegible]

REFERENCES

ARMSTRONG, C. N. (1953). *Proc. Roy Soc. Med* 46, 301.
ARMSTRONG, C. N. (1955a). *Brit. med J.* i, 1173.
ARMSTRONG, C. N. (1955b). *Lancet*, i, 1051.
ARMSTRONG, C. N. (1958). *Proc. Roy. Soc Med.* 51, 23.
ARMSTRONG, C. N. (1958). *Symposium on Nuclear Sex.* Interscience Pub Inc. p. 84.
ARMSTRONG, C. N., GRAY, J. E., RACE, R R. and THOMPSON, R. B. (1957). *Brit. med. J.* ii, 605.

V

Radiation as a Toxic Agent

R. H. MOLE

It is now generally recognized that ionizing radiation can do harm and there is often an unexpressed feeling that there is something special or unique about the toxic action of radiation. Certainly no other human invention is as uniquely destructive as an atomic bomb but the fear and anxiety which radiation engenders seem to depend partly on the belief that we are conspicuously ignorant of what radiation does and how it does it, and partly on the belief that small doses of radiation are harmful, indeed that no dose, however small, is without some harmful effect. Yet there is a large body of knowledge, both clinical and experimental, of the toxicity of radiation and, although numerous powerful and potentially lethal chemical agents are in everyday medical use, no one seems to feel that these also are inevitably and always doing harm. Thus there is an interesting contrast between the toxic action of chemical agents and of radiation, in our mental approach to them at least (cf. World Health Organization, 1958), and it may be worth while to examine the salient features of biological damage by radiation from this comparative point of view.

Ease of Measurement of Radiation

In one way ionizing radiation must appear to be a mysterious agent since its presence cannot be detected with the unaided senses. But this is just as true of some chemical agents, like carbon monoxide, and of harmful levels of many toxic substances in the air we breathe or the food we eat, even though in higher concentrations they may be sensed. On the other hand

Wilkins, L., Gardner, L. I. and Criglie, J. F. (1952). *J. clin. Endocr.* **12**, 257.

Wilkins, L. and Migeon, C. J. (1951). *J. clin. Endocr.* **11**, i.

Wilkins, L., Lewis, R. A., Klein, R. and Rosenberg, E. (1950). *Bull. Johns Hopk. Hosp.* **86**, 249.

Witschi, E., Nelson, W. A. and Segal, S. J. (1957). *J. clin. Endocr.* **17**, 737.

quantitative phenomena as amount of drug given and sensitivity of recipient. In fact the common usage is excusable since with many drugs the most important factor is the sensitivity of the recipient, which is not very amenable to measurement: the individual who suffers a toxic reaction is usually the exceptional individual with a much greater sensitivity than the average. The situation is very different where ionizing radiation is concerned.

Absence of exceptional susceptibility to radiation

Firstly no one has ever described an innate idiosyncrasy or acquired hypersensitivity to radiation in the usual sense of reactivity to a single dose a hundred or a thousand times less than the dose required to affect an average individual. If such exceptional susceptibility did exist, occasional individuals would have shown marked toxic effects from the small radiation doses unavoidably received when undergoing routine radiological examinations for diagnostic purposes and the clinical experience of the past sixty years is enough to rule this out. (On this issue clinical experience is clearly of far more value than animal experiment.)

Widely differing radiosensitivity of different tissues

Secondly it is hardly possible to make a useful statement about the toxicity of radiation without bringing in the dose required to produce the effect under consideration. Thus it is perfectly true that radiation can cause sterility, cataract, cancer and even sudden death, but such a statement is really grossly misleading if it omits to add that the doses required to produce these various effects differ by up to a thousandfold. And, fortunately, since exceptionally susceptible individuals do not seem to exist, [illegible]

organs and functions is illustrated in Table 1.

The uniformity of the acute lethal response

One striking difference between ionizing radiation and chemical, including bacterial, toxins, is a remarkable uniformity in

the detection of ionizing radiation by physical means is now a straightforward matter. It is easy, and comparatively cheap, to get a measure of the radiation reaching any desired spot and to obtain, if desired, continuous records to show fluctuations or mean levels of radiation intensity over any desired period of time. This is an important property of radiation from the present point of view since the easy availability of radiation measurements positively stimulates, not only the making of more measurements, but also the asking of questions about their meaning, that is their correlation with biological damage of any kind. In these circumstances it should not be surprising that many of the possible questions as yet have no answer. In contrast, measurements of chemical concentration are nearly always much more complex and, unless they can be made by physical means, are much less adaptable to routine or automatic use. Thus questions about chemical toxicity are not as a rule provoked by the accumulation of chemical measurements, as questions about radiation toxicity are by the accumulation of radiation measurements. For example, it is only *after* geographical differences in cancer incidence have been revealed in epidemiological surveys that chemical measurements have been undertaken of the concentration of possible toxic agents in the air or soil of different places, whereas the *possibility* of geographical differences in the incidence of leukaemia or of genetic mutations is currently being investigated because geographical differences in background radiation levels have suggested the investigation.

Quantitative Aspects of Radiation Toxicity

Another consequence of the relative difficulty of making chemical measurements is that we tend to say that some chemical agent is a cause of some particular kinds of harm without reference to the amount or concentration of agent required. It might be said, for example, that the causes of agranulocytosis include the sulphonamide drugs, thiourea and its derivatives, amidopyrine, organic arsenicals and chloramphenicol, and the form of statement leads one to think of causation in qualitative, not in quantitative terms. But clearly the harm that may follow the exhibition of these chemical agents depends on such

Specific treatment critically dependent on dose

Large doses of whole-body irradiation may kill because the bone marrow is depleted of too many cells and such doses may also markedly depress the immune reactions of the body, especially the reaction to a new antigen not previously encountered. Thus the possibility arises of preventing the fatal outcome by substitution therapy, by providing the needed marrow cells from some other healthy individual (Loutit, 1956). The irradiated recipient may be a special case and an exception to the general rule that in man tissue transplantation between individuals is not successful because of the inevitable genetic and therefore antigenic differences between donor and recipient. Treatment with living cells from the bone marrow or from the haemopoietic foetal liver may be specific treatment for exposure to large doses of whole body irradiation.

However, treatment, as well as prognosis, may depend critically on the estimate of the radiation dose. If the dose was four-fifths or less of an LD_{50}, expectant treatment is probably all that is required, and an attempt to graft bone marrow cells from a donor would be contraindicated since experimental work suggests that this might then be positively harmful (Table 2). If the

TABLE 2 The effects of specific treatment and of size of radiation dose on the mortality of CBA mice after whole body irradiation (data of Trentin, 1956)

	Mortality	
Dose of radiation	Not treated	Injected intravenously with a suspension of cells from bone marrow*
330 r	nil	nil
550 r	12%	100%
660 r	90%	17%
770 r	100%	nil

* The marrow was from an F_1 hybrid, Cb × CBA

dose of radiation was larger, by as little as half as much again or more, the prognosis is so bad that many would consider that active treatment with suspensions of marrow cells should be instituted as soon as possible. The same relatively small increase

the acute lethal response to whole-body irradiation. With chemical agents species differences are to be expected in the dose required to kill, sometimes of up to a thousandfold, and within one species there will be quite a range in sensitivity

TABLE 1. Differential tissue sensitivity to single doses of X or gamma radiation in the adult experimental animal: illustrative examples

Dose rads	
25–50	Marked impairment of fertility in the female mouse
	Definite but recoverable damage to the spermatogenic elements of the testis
	Temporary marked reduction in normoblast concentration in the bone marrow
	Immediate (2–24 hours) death of some small lymphocytes
	Delay in emptying time of the stomach
	Accelerated onset and increase in incidence of the ovarian tumours characteristic of old age in the mouse
	Gene (point) mutation rate doubled
500	Acute radiation death within a few weeks after whole body ex- [illegible]
1000	[illegible]
2000–3000	Marked vascular damage, local necrosis of tissue with permanently unstable scarring
	Radiotherapeutic range
100,000	Sudden death within a few hours

There may be marked variations in radiosensitivity during the period of development from zygote to mature adult, and in the foetus, infant or child differential tissue sensitivity, as well as absolute sensitivity, may be very different from the adult.

amongst different individuals even if they are of the same sex, age and weight. With ionizing radiation, however, the LD_{50} for an unusually wide variety of mammals has been determined, and the largest value is not more than about four times the smallest (Bond and Robertson, 1957). There is a correspondingly small difference in sensitivity between different individuals, so that a 20 per cent increase in dose might well increase the probability of dying after an acute whole-body exposure from 10 per cent to 50 per cent.

general the slow evolution of radiation damage is even more characteristic of damage to humans than of damage to animals.

The testis has now been the subject of radiobiological investigation for over fifty years and the pattern of cellular changes in the irradiated testis appears to be very simply explicable. The radiation interferes with the process by which the numbers of

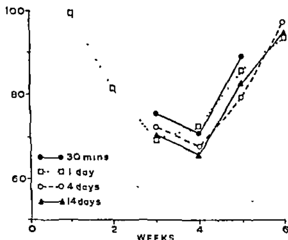


Fig. [illegible] The weight of the testis of CBA [illegible]
Abs [illegible]
Ora [illegible]
am [illegible]
in the figure.

spermatogonia are kept constant, both by killing spermatogonia directly and by preventing successful cell division of the primitive stem cells. The reduction in the numbers of spermatogonia has no immediate effect on the mass of the testis because these cells form such a small proportion of the whole organ. However, reducing the number of stem cells must lead to a reduction in the number of their progeny and later on there will be a smaller number of [illegible]

sperm count in an ejaculate will be decreased. The time taken from the moment of irradiation to the point where the sperm

in dose of radiation which worsens the prognosis so drastically that specific treatment is needed, also makes this specific treatment possible, by reducing the immune responsivity of the body below the level at which harmful reactions to the injected foreign cells will occur, thus allowing the grafted marrow cells to survive and multiply within the irradiated host, to its benefit rather than to its harm. The exact dose of radiation received is thus of paramount importance, since it determines whether specific treatment—by injection of marrow cells—will do good or harm. There seems to be no parallel amongst other kinds of toxic agent.

The estimation of the radiation dose received in accidental exposures is likely always to be a matter of considerable difficulty and complexity and the health physicist's task is made no easier by the speed and accuracy demanded of his calculations, and the knowledge of their overriding importance to the patient.

Time and the Assessment of Radiation Damage

As important as dose in assessing radiation damage is time, for special features of radiation damage are its slow rate of evolution and the way in which spreading a dose in time affects the amount of damage.

The testis

The testis is one of the most radiosensitive organs. The change in weight of the mouse testis after a single small dose (5–10 per cent of the acute lethal dose) is shown in Figure 1. There is no change during the first week after which the weight begins to fall rapidly reaching a minimum at three to four weeks. There is then a loss of about half of the sperm-producing moiety of the organ. Recovery appears complete in another two to three weeks.

The same pattern of change has been seen in the sperm counts of men exposed to doses of a similar magnitude (Hasterlik and Marinelli, 1956). The delay in reaching the period of maximum damage was even more obvious: the time scale in the human was greatly extended, 7 to 10 months being needed to reveal the full effect of the small momentary exposure. In

TABLE 3. Average number of young born to female C57 Bl mice during different periods after a single dose of radiation

Dose of X-rays	Weeks after irradiation 0–12	13–24	25–36	37–48	49–
Control	11	17	10	5	2
25 r	11	10	1	0	0

The mice were intensively mated in order to measure the 'functional reserve' of their reproductive ability (Mole, 1959b).

Two additional points may be made. First, the dose of radiation used was of the order of dose which geneticists feel is important for genetic reasons.[1] It should not be overlooked that such doses may also have important somatic consequences which are highly relevant to the interpretation of genetic experiments designed to measure the mutation rate per unit of radiation dose (Mole, 1959b). Secondly, it is commonly taught that even as early in life as at birth the mammalian ovary contains all the primordial oocytes it ever will have and has already lost all its oogonia. Without these stem cells new oocytes cannot be produced. Thus a loss of oocytes cannot be compensated for and radiation damage to the ovary is permanent and irrecoverable unlike radiation damage to the testis. There is no essential difference in the kind of radiation damage produced by radiation in ovary and in testis: in each organ radiation specifically reduces the numbers of the least mature cell types, primordial oocytes and spermatogonia respectively. The difference in outcome depends on the intrinsic biological differences between the two kinds of cell, one of which can reproduce itself, one of which cannot.

THE OUTCOME OF TOXIC DAMAGE DEPENDS PRIMARILY ON THE INTRINSIC POTENTIALITIES OF THE AFFECTED TISSUE

The comparison of testis and ovary illustrates an important general principle. The outcome of toxic damage to a tissue does not depend primarily on particular properties of the toxic agent but rather on the ways in which cells can respond to it, and

[1] The human ovary may be rather less radiosensitive than the ovary of the mouse

count is least will be a measure of the time taken for the normal physiological sequence of cell divisions from stem cell to finished product. Recovery in numbers of spermatogonia is already well under way at the time the number of spermatozoa is minimal, but again time is needed for the newly arising stem cells to undergo the successive divisions and maturations necessary to restore the normal population of cells in the organ. Hence the delay in recovery.

The same general picture will be found after irradiation in any tissue, such as bone marrow, intestinal epithelium or skin, which normally loses its most mature and differentiated cells and normally continually replenishes itself by an ordered sequence of cell divisions starting from primitive stem cells. Such tissues seem always capable of recovery if sufficient time is available and if nutrition is not interfered with by vascular damage or infection. The time required for recovery is closely dependent on the normal renewal time of the tissue concerned so that recovery of the intestine is faster than the recovery of the bone marrow and recovery of the testis is slowest of all.

Figure 1 also shows the change in weight of the mouse testis after a fixed dose of radiation when the period of time over which the radiation was given was varied nearly seven hundred-fold, from half an hour to a fortnight. Spreading the dose in time, i.e. decreasing the rate at which the dose was given, did not decrease the effect, quite unlike what would be expected from spreading the dose of a chemical agent in time. (Spreading the dose still further in time does change the response of the testis as discussed later.)

The ovary

A small dose of radiation, such as 25 r, may also have a marked effect on the ovary (Table 3). Again time was needed for the full extent of the damage to fertility to become visible; no effect at all was apparent during the first three months after irradiation. Only after the first quarter of the mouse's reproductive life span had elapsed did the reduction in fertility begin. Yet the degree of damage was not negligible: the overall reproductive capacity was halved.

Delayed Lethal Effects

Apart from damage to the gonads, the practical aspects of chronic radiation toxicity centre around the delayed killing effects of radiation. Acute aplastic anaemia should no longer be seen in radiation workers since the level of radiation dose required to cause a physiological deficiency in the bone marrow is far higher than the safety levels of ten years ago which in their turn are several times higher than present-day maximum permissible levels. I do not know of any published record during the past twenty years of aplastic anaemia due to occupational radiation exposure.

The delayed killing effects of radiation which may be important are malignant disease, including leukaemia, and accelerated or premature ageing. There is good evidence that malignant disease of several different kinds has been produced in man by irradiation, and perhaps at present the greatest interest is centred on the question of how much radiation is required to produce cancer or leukaemia. As for premature ageing, it has been claimed that American radiologists died at an earlier age than other medical men, and the evidence has often been cited (as in the United Nations Report, 1958) as showing that occupational exposure shortens life in some other way than producing cancer. Experimental work has also been interpreted as meaning that radiation can cause accelerated or premature ageing, and, although in my view the interpretation is probably mistaken, the idea has stimulated a great deal of interest and a proliferation of hypotheses which there is no space to discuss here The clinical evidence seems clear. British radiologists lived longer, if anything, than other British medical men (Brown and Doll, 1958) and a re-analysis of the American data has shown that there never was any real evidence of life shortening in American radiologists either (Seltser and Sartwell, 1958). The incidence of leukaemia in American and skin cancer in British radiologists was certainly increased so that the induction of malignant disease seems a much more important practical problem than non-specific life [illegible]

there is only a limited number of ways in which any one kind of cell can respond. The cells of the central nervous system, for example, have only one way of responding to anoxia and this is independent of the particular agent responsible for the cerebral anoxia. Similarly the consequences of radiation damage depend on the intrinsic potentialities of the affected tissues and are not likely to be qualitatively unique to radiation in any fundamental sense.

The complexity and pattern, or specificity, of pathological reactions is commonly a consequence of the interaction of many different kinds of cell. It is characteristic of radiation damage by doses in the range which concern those interested in environmental or occupational hazards that this interaction is usually absent. The underlying cause of the lack of interaction of different cell types is the wide range of radiosensitivity so that in any one tissue only one kind of cell may be affected: in the testis, for example, spermatogonia but not spermatocytes, spermatids, spermatozoa, Sertoli cells, or the cells of interstitial, vascular or connective tissue. The characteristic lesion of radiation damage is simple cell depletion. Tissue changes caused by radiation are thus non-specific unless the dose is up in the radiotherapeutic range, when a variety of different cell types is affected, cellular interaction becomes possible and aetiologically specific tissue reactions may become recognizable, as in the lung (Whitfield, Bond and Arnott, 1956) or the kidney (Luxton, 1953).

Adaptation to Chronic Irradiation

The difference between tissues which can reconstitute themselves from their own stem cells and tissues which cannot is of great importance to the assessment of damage from long-continued daily irradiation. With a small daily dose of radiation the daily loss of cells may be small enough for it to be compensated for by an increased number of cell divisions, leaving the cell population effectively unchanged. This is what happens in the chronically irradiated testis and a similar compensation or adaptation to chronic irradiation has been observed in the peripheral blood count and in other indices of damage (cf. Mole, 1959a).

fractions was increased to 20 and more. In view of the possibility of other sudden changes in the kind of biological response at other points in the scale of fractionation, it would not be wise to extrapolate beyond the data and say that 1,000 fractions would be more harmful than 100, especially as there was no evidence of a progressive increase in effect as the number of

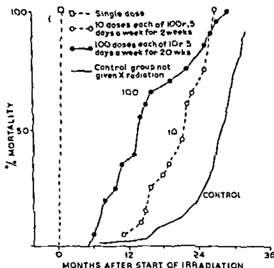


FIG 2. Mortality of groups of female CBA mice each given a total dose of 1000 r of X-rays but in a different number of fractions to each group (Mole, 1959c).

fractions increased from 20 to 100. It is not possible with experimental animals like the mouse to give 1,000 r in 1,000 daily doses and still leave time for the effects of radiation to develop. The animals do not live long enough.

With the smaller total dose of 750 r there was a similar increase in damage and a similar change in the biological response to the radiation when the duration of the exposure was increased (Mole, 1959a). In other experiments on mice it was found that when a protracted course of daily irradiation sufficient of itself to produce a high incidence of leukaemia was followed by an additional large single dose, the incidence of leukaemia was

only when the killing disease is naturally uncommon. In experimental work, however, the number of individuals in a group is never large enough for a real increase in some killing disease to be compatible with an unaltered life span. In fact the average life span may be closely dependent on the frequency of occurrence of some particular type of malignant disease and this must always be borne in mind when considering the interpretation of experimental results on shortening of life by irradiation.

Effects of protraction and fractionation

With mechanical and chemical injury we commonly expect that the extent of an injury will be progressively decreased by increasing the fractionation of the injurious agency and the time over which it acts. The loss of two litres of blood in two minutes is likely to be lethal; the same loss spread out over two days will produce symptoms and probably lead the victim to seek medical care. Yet if the two litres are lost over two years the loss might well be difficult even to detect and would certainly cause no harm. The same principle underlies all drug therapy; the effect of any one dose wears off and successive doses need to be given to maintain an effect. And the principle is also true of the acutely killing effects of radiation: spreading the dose out in time increases the dose needed to kill.

But with delayed effects of irradiation, unlike the acutely killing effects, this may be far less true: in fact spreading a dose in time may even increase its killing effects. In these experiments the overall dose of radiation was given in times ranging from a few minutes to 4½ months. The animals were then kept for the rest of their lives free from added irradiation. Every opportunity was thus given for any delayed effects to become manifest. As can be seen (Fig. 2) a given total dose of 1,000 r killed mice earlier when it was given in 100 daily fractions (irradiation period 20 weeks) than when it was given in 10 successive daily fractions (irradiation period a fortnight). This seemed to be because of a change in the kind of biological response as the fractionation of the dose or the time over which it was given increased. The incidence of leukaemia and reticulosarcomata increased markedly when the number of daily

work that point mutations are produced in strict proportionality to the dose of radiation, that the time over which the radiation is given and the rate of its administration are immaterial and that there is no threshold dose below which radiation has no effect. The same relations should hold for point mutations in somatic cells, and work with radiation therefore provides a *means of testing the hypothesis that point mutation is the cause* of malignant disease. As will be seen the point mutation hypothesis stands up to test very poorly, while on the other hand direct observation suggests that directly visible chromosomal mutation may be of fundamental importance. The two sorts of mutation lead to quite different conclusions about the amount and kind of irradiation required to cause cancer so that the discussion has a practical as well as an academic interest.

Point mutations and dose intensity

Experiments cited earlier showed that the delayed effects of a given total dose of radiation varied markedly with the way the dose was spread in time. This of itself is not easily compatible

TABLE 4. Leukaemia incidence in female mice after a four-week exposure to ^{60}Co gamma radiation (Mole, 1959c)

Irradiation schedule	Daily dose r	Dose rate r/hour	Total dose r	Leukaemia 15 months after start of irradiation CBA*	C57Bl*	combined
5 days/week	50	81	1000	12/30	11/29	39%
5 days/week	50	21	1000	8/30	13/30	35%
5 nights/week	50	3 3	1000	7/25	3/29	19%
Continuously	32	1 3	900	2/30	1/25	5%
Controls	nil	nil	nil	0/30	1/28	2%

* Number of mice with leukaemia/number in group at start of experiment

with the point mutation hypothesis, but a much more stringent test is provided by an experiment in which a fixed total dose of radiation was given in a fixed overall period of time but where the dose intensities of the individual radiation exposures varied between the different groups. The results (Table 4) show that the higher the dose intensity the greater the incidence of leukaemia. Such a result may be thought to amount to a formal

markedly reduced (Mole, 1956). These may all be particularly dramatic examples but it is clear that the amount of damage produced by a dose of radiation, whether it is measured by degree of life shortening or by the incidence of particular kinds of tumour, depends not only on total dose but to a very marked extent on just how the radiation is given, i.e. the circumstances of the irradiation. No theory which assumes that damage depends simply on dose can possibly be true. Nevertheless the dose levels used in these experiments, although often below the radiotherapeutic range, may still have been large enough to lead to cellular interaction and imperfect repair: with smaller doses of radiation the relation between dose and damage may be different.

Carcinogenesis

One feature of radiation damage already emphasized is the way in which time is required for overt clinical harm to develop; it has been said that in a sense living things have a memory of being irradiated. A long latent period, however, is characteristic of carcinogenesis as a process, not merely of radiation, and any carcinogenic insult must be imagined to leave some permanent imprint behind it. This permanent change is often taken to be a change in the genes or chromosomes of affected cells. The idea of somatic mutation in some form or other enters into many hypotheses of carcinogenesis, and is certainly not incompatible with theories that cancer is induced by infective agents.

Somatic mutation

[illegible]

mutation on the other hand is a change in a single gene and thus its occurrence cannot be directly observed but only inferred: genes and point mutations are 'theoretical entities'. In work with sex cells inferences may be made from changes in Mendelian inheritance but this sort of evidence cannot be derived from somatic cells since they do not undergo sexual conjugation. It has been generally accepted, however, in genetic

1958). The causal importance of these visible chromosomal changes must await further investigation but it is suggestive that the production of some kinds of chromosomal mutation is known to depend critically on dose intensity, just as leukaemia induction seems to.

The possibility of a threshold for the induction of cancer by radiation

If carcinogenesis depends merely on a chromosomal mutation a certain probability of inducing cancer would be expected to follow any radiation exposure, however small, even though the exact probability would depend on circumstances; the lower the dose intensity, for example, the lower the probability. Even without a simple linear proportionality of dose and cancer there might well be no threshold dose. However it is increasingly believed that more than one step is required for the induction of malignant disease and radiation carcinogenesis may depend on more than just chromosomal mutation. The common antecedent of malignant disease when it occurs as a late complication of therapeutic irradiation is severe tissue damage with scarring. Skin cancer as an occupational hazard to radiologists has certainly been greatly reduced, and perhaps eliminated, since radiation exposure has been reduced below the levels which produce marked skin damage. Clinical experience therefore suggests that the probability of devel [illegible]

with this conclusion. Irradiation can result in a high incidence of tumours in a variety of tissues (cf. Mole, 1958) and whenever the irradiated tissues have been examined, gross tissue damage has been found to precede the development of these tumours. In these situations the relationship of dose to tumour has been markedly curvilinear, the increase in tumours following an increase in dose being much greater than would be expected on a simple linear hypothesis.

These observations, though perhaps suggestive, do not prove that a minimal degree of tissue damage is an essential prerequisite for the development of a tumour and therefore that there is a threshold of radiation exposure below which cancer

disproof of the hypothesis of point mutation for murine leukaemia at least.

Murine leukaemia is well known to differ in a number of ways from human leukaemia, but the differences are not important in this context since murine leukaemia is undoubtedly a malignant disease, and it is the general hypothesis that point mutations are responsible for carcinogenesis which has been tested and found wanting. Moreover it should be emphasized that the evidence that radiation can cause leukaemia in man (*British Medical Journal*, 1959) also comes from situations where the dose intensity of individual exposures was high, as during exposure to an atomic bomb, to therapeutic irradiation or to diagnostic irradiation. (The exposure of radiologists to radiation in the course of their work will have been in either or both of the latter two categories.) If the experimental evidence for the importance of dose intensity in leukaemogenesis is accepted, it is clear that it is not legitimate to use the observed human data relating dose and leukaemia incidence to deduce how much leukaemia would be produced by an occupational radiation exposure or an increase in level of background radiation. The average dose intensity of background radiation is at least 100 million times less than the dose intensities of the radiation exposures known to have caused leukaemia in man.

Unbalanced chromosomal changes and cancer

There have been many attempts to define a fundamental difference between normal and neoplastic cells in biochemical or morphological terms but so far no distinction has been found to be universally valid: there have always been exceptions. It is of great interest therefore that direct cytological observations in murine leukaemia, both spontaneous and radiation-induced, have shown that characteristic changes in the chromosomal complement are the rule (Ford, Hamerton and Mole, 1958). There are three kinds of changes, in modal number, in distribution around the mode, and in morphology. For technical reasons observations in human leukaemia are much less easy to make but preliminary results suggest that similar changes may also be found in human neoplastic cells also (Ford, Jacobs and Lajtha,

the presence or absence of a threshold dose of a toxic agent may be proved in any absolute sense. This is as true of chemical agents as of radiation and it would be an interesting, though not a purely scientific, enquiry to ask why there has been intensive discussion of the possibility of a threshold for the toxic action of radiation rather than, say, of aspirin or caffeine. There is, in

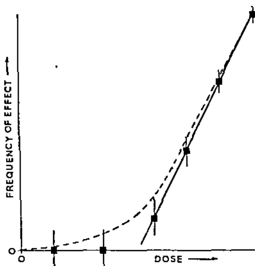


FIG 3 Schematic diagram relating dose of radiation and frequency of effect to illustrate the logical impossibility of proving the existence of a threshold from observational data alone (see text)

fact, very little experimental information on the truly chronic toxicity of chemical agents of any kind, and in practice the question of an experimental demonstration of a threshold for the toxic action of chemical agents is disregarded. On the one hand drugs with known toxicities continue to be used, on the other hand substances are banned altogether from use as food preservatives or additives after relatively large concentrations have been found to be deleterious to experimental animals in relatively short-term and sometimes very artificial experiments.

will not be produced. What sort of evidence could there be on this question? The best-documented cases of human radiation-induced leukaemia, in Japanese survivors of atomic-bomb explosions and in ankylosing spondylitics given therapeutic irradiation, are of little help because the radiation exposures resulting in a real increase in leukaemia were large enough to [illegible] rts [illegible] ia [illegible] his context. For this and other reasons confirmation of the findings is urgently needed, and in addition some analysis of the radiation doses received in relation to the amount of cancer produced. Radiation doses in diagnostic radiology have not always been too small to produce tissue damage (cf. Table 1).

The Idea of a Threshold for Toxic Action

Observational data alone will never be sufficient to decide whether there is or is not a threshold for radiation carcinogenesis: the decision will have to be made on purely theoretical grounds (Mole, 1958). Suppose observations are made as in Figure 3 which suggest that there really is a threshold. It must be accepted that each observed point has some statistical uncertainty. It is thus open to anyone to suggest that the real relation between dose and effect is that shown by the dotted line in Figure 3. In fact the lower the observed frequency of effect, the smaller the number of observed cases and the greater the statistical uncertainty, so that it is logically inevitable that the least securely established points will be the points on the lower left of the diagram, the points which might be thought to be the most important from the point of view of establishing the existence or not of a threshold. The argument is a purely logical one and applies to all varieties of toxic action, not only carcinogenesis.

The existence or not of a threshold dose below which no effect occurs is a theoretical question which depends on the particular hypothesis chosen to account for the phenomena under examination. Since there is no limit to human ingenuity and therefore to the invention of theories, there is really no criterion by which

the presence or absence of a threshold dose of a toxic agent may be proved in any absolute sense. This is as true of chemical agents as of radiation and it would be an interesting, though not a purely scientific, enquiry to ask why there has been intensive discussion of the possibility of a threshold for the toxic action of radiation rather than, say, of aspirin or caffeine. There is, in

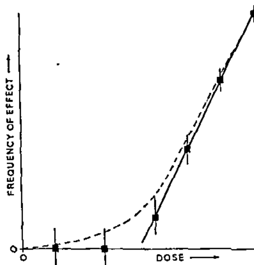


FIG 3 Schematic diagram relating dose of radiation and frequency of effect to illustrate the logical impossibility of proving the existence of a threshold from observational data alone (see text).

fact, very little experimental information on the truly chronic toxicity of chemical agents of any kind, and in practice the question of an experimental demonstration of a threshold for the toxic action of chemical agents is disregarded. On the one hand drugs with known toxicities continue to be used, on the other hand substances are banned altogether from use as food preservatives or additives after relatively large concentrations have been found to be deleterious to experimental animals in relatively short-term and sometimes very artificial experiments.

Basic Mechanisms of Radiation Toxicity

Various aspects of radiation toxicity in the whole animal have been considered without any reference to mechanisms by which radiation kills cells or interferes with the process of cell division. Most accounts of radiobiology lay great stress on these mechanisms but in my view they are not as important in considering radiation toxicity as they are in radiotherapy. What is important to the whole animal, especially when considering the delayed toxic effects of radiation, is not so much the cells which are killed as the cells which survive, and especially the cells which retain the power of cell division. It is these which multiply and replace those which have been lost, thus restoring function to normal, and it is clearly from surviving cells that neoplasia must originate.

A most interesting academic question is whether a stimulus to cell division over and above the normal rate can have any harmful consequences of itself. For example, if the hypothesis is accepted that cancer is due to chromosomal change and if the further assumption is made that such a change has a certain probability of occurring spontaneously at each cell division, then an increase in the rate of cell division must lead to an increase in the incidence of cancer. Alternatively, is it possible to exhaust the capacity of stem cells to divide so that a sudden tissue failure supervenes? Thus, it is quite characteristic of both men and animals exposed to dangerously high levels of daily irradiation that the terminal fatal anaemia sets in quite suddenly after a long period during which there is little significant anaemia (Mole, 1954; Edmondson, 1958). However, the experimental observations make it unlikely that this is because the capacity of the stem cells of the bone marrow has been suddenly exhausted, since at death the marrow is full of normoblasts, and [illegible] of the anaemia seems to be due to the progres- [illegible] critical low level. Although the idea of damage or strain of a normal organ from overuse is part of the folklore of medicine, it seems to be supported by little positive evidence.

The long-term survival of irradiated somatic cells was until

recently rather a neglected subject, and another important problem is whether cells which have survived irradiation are fully normal. As well as the unbalanced chromosomal changes which seem to be characteristic of neoplasia, as already discussed, balanced chromosomal changes also occur which are not related to neoplasia but which may be responsible for physiological deficiencies in cells which superficially might appear normal (Barnes, Ford, Gray and Loutit, 1959). The full significance of these chromosomal changes is certainly not yet appreciated since they have been detected only as a result of recent technical developments in mammalian cytogenetics. Invisible point mutations in the genes of somatic cells must also occur, but their demonstration must await techniques which have yet to be worked out. Such genetic changes, large or small, might be important in somatic stem cells. They could lead to a change in the natural balance between the different functional properties of a cell clone so that there might result, for example, a race of stem cells with a higher potential for cell division than the unchanged natural line but with a lower ability to produce functionally useful, mature end-product cells, cf. the uncommon cases of aplastic anaemia with an apparently active marrow. Such speculation can be intensely interesting but is still merely speculation: that consequences of this kind can follow exposure to radiation has still to be established.

REFERENCES

Barnes, D. W. H., Ford, C. E., Gray, S. M. and Loutit, J. F. (1959). *Progress in Nuclear Energy*, Series VI, Biological Sciences, 2, 1 Pergamon Press, London.

Bond, V. P. and Robertson, J. S. (1957). *Ann. Rev. Nuclear Science*, 7, 135

British Medical Journal (1959). i, 1095.

Brown, W. M. C. and Doll, R. (1958). *Brit. med. J.* ii, 181

Edmondson, P. W. (1958). *Brit. med. J.* i, 363.

Ford, C. E., Hamerton, J. L. and Mole, R. H. (1958). *J. Cell. comp. Physiol.* 52, Supplement 1, 235.

Ford, C. E., Jacobs, P. A. and Lajtha, L. G. (1958). *Nature*, 181, 1565.

Hasterlik, R. J. and Marinelli, L. D. (1956) *Peaceful Uses of Atomic Energy*, 11, 25, United Nations, New York.

LOUTIT, J. F. (1956). *Lectures on the Scientific Basis of Medicine.* Athlone Press, London, v, 439

LUXTON, R. W. (1953). *Quart. J. Med.* n.s. 22, 215.

[illegible]

[illegible], 1955, ed. Mitchell, Holmes

[illegible]

[illegible]

SELTSER, R. and SARTWELL, P. E. (1958). *J. Amer. med. Assoc.* 166, 585.

TRENTIN, J. J. (1956). *Proc. Soc. exp. Biol. Med. N.Y.* 93, 98.

UNITED NATIONS SCIENTIFIC COMMITTEE ON THE EFFECTS OF ATOMIC RADIATION, Report of (1958). General Assembly, Official Records: Thirteenth Session Supplement no. 17 (A/3838), New York.

WHITFIELD, A. G. W., BOND, W. H. and ARNOTT, W. M. (1956). *Quart. J. Med.* n s. 25, 67.

WORLD HEALTH ORGANIZATION (1958). Mental Health Aspects of the Peaceful Uses of Atomic Energy, Technical Report Series 151.

VI

What are Gamma Globulins?

J. H. HUMPHREY

THE problem of the nature and origin of gamma globulins has become more insistent during the last few years, as an increasing number of techniques has become available for their study. However, already in 1954, Grabar, whose work has contributed greatly to our knowledge of this subject, gave a lecture with the same title as this, and I have inevitably borrowed a number of the ideas contained in it. The discussion given below is intended to be comprehensive, and possibly provocative, but should not be taken as a complete review of the subject.

The term gamma globulin is derived from electrophoretic analysis—the definition put forward by the U.S. Commission on Plasma Fractionation and Related Processes being that gamma globulins are the fraction of plasma which does not contain proteins with electrophoretic mobility greater than $-2{\cdot}8 \times 10^{-5}$ in veronal buffer pH 8 6, ionic strength $=0$ 1. This serves well enough as a working definition—it allows recognition of a group of proteins with generally similar properties, both biological and physico-chemical, which are present in the plasma of all mammals examined, and in at least some fish and a number of reptiles, but are absent from arthropods, arachnidae and molluscs (Deutsch and Goodloe, 1945; Dessauer and Fox, 1956; Woods, Paulsen, Engle and Pert, 1958). But difficulties arise when an attempt is made to relate antibodies, natural or immune, to each other and to so-called normal gamma globulin and to study the distribution of proteins which are antigenically related to gamma globulin in the different plasma fractions.

Similar difficulties arise in considering pathological proteins found in cases of myeloma and of Waldenstrom's macroglobulinaemia.

It is now generally agreed that gamma globulins are complex, and how complex they are is becoming increasingly apparent.

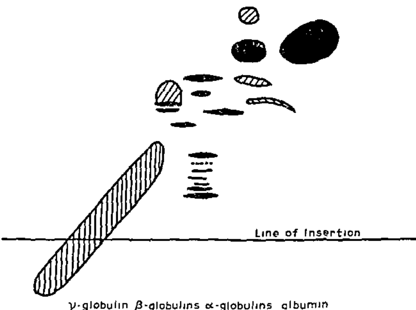


FIG 1 Diagram of two-dimensional starch-gel electrophoresis [redrawn from Smithies (1955)], illustrating difference in behaviour of γ-globulin from that of other serum proteins.

This can be illustrated by referring to three separate techniques which have been used, incidentally, to study gamma globulins. The first is two-dimensional starch gel electrophoresis (Smithies, 1955). Figure 1 shows that gamma globulin from Man (or any other species) moves as a long spread-out diagonal line, indicating a family of proteins with closely similar overlapping electrophoretic mobilities and sizes. The contrast with other well-defined proteins—e.g. iron binding protein, haptoglobin, albumin—is obvious. The second is immunoelectrophoresis, introduced by Williams and Grabar (1955). Here serum, or a gamma globulin preparation, is moved electrophoretically in agar; anti-

serum against one or more of the constituent proteins is allowed to diffuse towards the proteins; and where the antibody and antigen interact there appear lines of precipitation. These lines are interpreted as follows: lines which are distinct or cross belong to unrelated antigens, whereas continuous lines indicate the presence of the same antigen, and lines which fuse partially indicate antigens which share some but not all their antigenic components. It is evident from Figure 2 that protein reacting

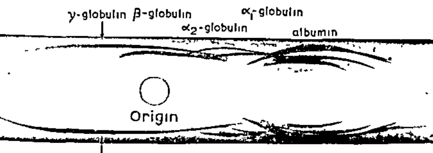


FIG. 2. Immunoelectrophoresis of normal rat serum. The serum was applied at O, and two different antisera were allowed to diffuse in at the sides. Arrows indicate gamma globulin lines.

with antibody against gamma globulin occurs in a long trail extending well into the region in which beta and even alpha globulins are found. Since the endosmotic flow in agar is rather great, the gamma globulins move towards the cathode and beta globulins remain around the origin.

The third technique is a rather different one—namely chromatography on a weakly basis ion exchange powder (DEAE cellulose developed by Peterson and Sober, 1956) using a gradient elution with gradually increasing salt and hydrogen ion concentrations. Figure 3, taken from a paper by Fahey, McCoy and Goulian (1958) illustrates the manner in which human gamma globulin is distributed in the eluate from such a column. A rather similar picture is given by rabbit, rat, and guinea-pig sera—and probably many others. It must be emphasized that this distribution is not arbitrary—Dr. P. Farthing and I (unpublished) have found that if a fraction is taken out, labelled with ^{131}I, added back to another lot of the same serum,

and re-run, it will appear in the same position. Furthermore, the electrophoretic mobilities of the gamma globulins which emerge early are the lowest, and they increase progressively in

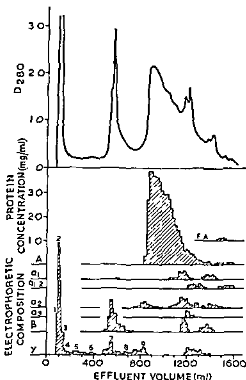


FIG 3 Elution diagram of normal human serum from DEAE cellulose column. Upper part of figure shows total protein concentrations. Lower part *shows composition of eluates as judged by paper electrophoresis. (From* Fahey, McCoy and Goulian, 1958)

successive fractions. The distribution of antibodies within these fractions will be referred to later.

These illustrations suffice to make the point that normal gamma globulin is complex and heterogeneous, but they by no means exhaust the varieties of heterogeneity. The molecular *weight of purified gamma globulin in man, and several other* species has been shown by ultracentrifuge and diffusion studies to be about 160,000, and the sedimentation constant is 6–7

Svedberg units (Kabat, 1939). However, it has been known for nearly twenty years that immune sera of ungulates contain a considerable proportion of molecules of molecular weight about 900,000 and sedimentation constant 18–19 Svedberg units—associated mainly, though not exclusively, with antibodies against carbohydrate antigens. These have been given various names—as is shown in Table 1, taken from a recent review by

TABLE 1. Various designations of the antibody-bearing globulin of hyperimmune horse sera of mobilities in the vicinity of $-2{\cdot}5 \times 10^{-5}$ cm ^{2} v.$^{-1}$ sec.$^{-1}$ in slightly alkaline buffer

Horse antibody against	Antibody designation	$-u \times 10^5$ cm ^{2} v.$^{-1}$ sec.$^{-1}$	pH
Pneumococcus	'Between γ and β'	2 1	8 0
Diphtheria	Pseudoglobulin	2 6	7·35
Tetanus	T-globulin	2 0	7·6
Diphtheria	β-globulin	2·3	8 0
Diphtheria and tetanus	T-globulin	2·2	7 6
Tetanus	T-globulin	1 8–2 6	8·5
Pneumococcus	AB-globulin	2·4	8 03
Diphtheria	γ_1-globulin	2 5	8·6
Tetanus	β_2-globulin	2·35	8 6
Diphtheria	β_2-globulin	2 35	8 6, 8 0

Taken from Isliker (1958)

Isliker (1958)—but they all refer to the same thing. Furthermore these fractions not only contain antibody, but they cross-react immunologically to an extent of about 60 per cent with the gamma globulins of molecular weight 160,000 in the corresponding sera (Treffers, 1944). Increased use of electrophoretic and ultracentrifuge techniques has revealed that normal human and rabbit sera also contain small amounts (2–3 per cent of total protein) of molecules with molecular weights around 900,000 which migrate electrophoretically as fast [illegible]

Kunkel, 1956). Their relationships to antibodies will be discussed later, but I will first discuss briefly a group of fascinating proteins, which must be taken into account in any consideration of the nature of gamma globulins—namely the pathological

proteins of patients with multiple myeloma or with Waldenstrom's macroglobulinaemia. These have nearly all been described in Man, but they may occur in other animals, and a transplantable myeloma of mice produces a typical myeloma protein with properties of a gamma globulin (Nathans, Fahey and Potter, 1958). Myeloma proteins are usually, if not always, produced by plasma cells very like those which produce antibody. They are rather easily purified and even crystallized, though crystallization by no means implies immunological or physico-chemical homogeneity, and they occur with electrophoretic mobilities ranging from those of slow gamma globulin to beta globulin or even alpha$_2$ globulins. Their size is usually about that of normal gamma globulin—i.e. their sedimentation constant is 6·6S, or occasionally 9·5S. Immunologically they *cross-react with antisera against normal gamma globulin;* but they may differ antigenically from one another when examined by the agar gel diffusion method (Slater, Ward and Kunkel, 1955). Chemical analysis has shown them to have amino acid compositions not unlike that of normal gamma globulins, but they do not necessarily have the same N-terminal end group (Putnam, 1957). Thus normal human gamma globulin has N-terminal aspartic and glutamic acid, while myeloma proteins may have one or other or both of these N-terminal groups, and one such protein has been described with N-terminal leucine. Macroglobulins are in many respects similar, but are composed of larger molecules—usually heterogeneous—with sedimentation constants 18S and upwards. They also cross-react immunologically with antibodies prepared against human gamma globulin, although sometimes to a relatively small extent only (Korngold and van Leeuwen, 1957). Some workers claim that they can prepare antisera, by absorption with normal human globulin, which react exclusively with the macroglobu-

treatment with reagents containing sulphydryl groups the macroglobulins examined can be depolymerized to components with sedimentation constants about 7S, without change in their

immunological properties (Deutsch and Morton, 1958). Another characteristic of pathological macroglobulins is their relatively high carbohydrate content, but in this they do not seem to differ essentially from the macroglobulins of normal serum.

What is the relationship of all these proteins to antibody? To generalize from one species to another, and from one kind of

TABLE 2 Molecular weights (M) and frictional ratios (f/f_0) of antibodies including partial specific volumes (V_{20}) and sedimentation constants (S_{20})

Antigen	Origin of antibody	V_{20}	$S_{20} \times 10^{13}$	M × 10^{-3} corr.	f/f_0
Pneumococcus	Horse	0·715	19·3	920	2 0
	Pig	0 744	18 0	930	2·02
	Cow	0 725	18·1	910	1·98
	Rabbit	0·745	7 0	157	1·4
	Monkey	0 74	6·7	157	1 5
	Man	0·745	7·4	195	1·5
Diphtheria	Horse	0·736	7 2	165	1·4
	Normal human sera				
	γ-globulins	0 74	7·2	156	1 38
Blood group	Isoagglutinins		19 8		
agglutinogens	A_1O serum[b]	0 74	15·5	500	1 36
	A_1A_1 serum[b]	0 74	11	300	1·37
	O serum[b]	0 74	6 5	170	1·58

Taken from Isliker (1958)

antibody to another, or even from one stage of immunization to another is dangerous, but unavoidable. Let us first consider acquired antibodies evoked in response to known immunizing agents. Many antibodies, in a variety of species, are found electrophoretically in the slow gamma globulin fraction, and have a molecular weight about 160,000. However, in horses and cattle particularly, antibodies may occur predominantly in a fraction moving electrophoretically between beta and gamma globulins, and have a molecular weight nearer 1,000,000, as is shown in Table 2 taken from Isliker (1958). Antibodies against the same antigen may, however, occur in both fractions of the plasma simultaneously or in different fractions at different stages of immunization. It is well known that human gamma globulin

prepared by organic solvent fractionation procedures contains many kinds of antibody—this is the rationale for treatment with 'immune globulin'—and this fraction is composed largely of gamma globulin with sedimentation constant 6–7S. However, if the fractions obtained by chromatography of human serum on an anion exchange powder, as shown in Figure 3, are examined for various kinds of antibody, antibodies may be found throughout all the fractions containing gamma globulin—antibodies against certain antigens often appearing characteristically in a given serum in particular positions. Reaginic skin sensitizing antibodies appear in the middle fractions (Humphrey and Porter, 1957)—while some isoagglutinins appear right at the end. When serum of a rabbit hyperimmunized against ovalbumin or against type III pneumococcus is run in the same way a similar sort of pattern emerges, but antibodies against ovalbumin or the capsular polysaccharide are found in all the fractions containing gamma globulin—which seems to imply that there is no fundamental difference between the fractions as regards their capacity to be or not to be antibody against any given antigen. The more carefully work is done on trying to locate and identify certain special antibodies, the more it becomes apparent that antibodies can differ in their physicochemical properties as much as can the components of normal globulins which, though antigenically similar, vary widely in electrophoretic mobility and even in molecular weight. Thus in Man most workers agree that useful antibodies, and such antibodies as 'blocking' antibodies, occur in the slow gamma fraction and have sedimentation constants about 6·3S; so do warm incomplete Rh antibodies (Fudenberg and Kunkel, 1957). On the other hand, reaginic antibodies (Cann and Loveless, 1957; Kuhns, 1954; but see Sehon, Harter and Rose, 1956); anti A and B isoagglutinins (Faure, Fine, Saint-Paul, Eyquem and Grabar, 1955); cold haemagglutinins (Christenson, Dacie and Croucher, 1957); warm complete Rh antibodies (Faure, Fine, *Saint-Paul, Eyquem and Grabar,* 1955); the Wasserman factor (Davis, Moore, Kabat and Harris, 1945); the rheumatoid arthritis factor which reacts with normal gamma globulin (Franklin, Holman, Muller-Eberhard and Kunkel, 1957), if

this is really an antibody, have all been located in fast gamma or beta fractions. A number have been associated with heavy components, with sedimentation coefficients 16S (Wurmser and Filitti-Wurmser, 1957), 18–19S (Campbell, Sturgeon and Vinograd, 1955), or, in the case of the rheumatoid arthritis factor, 22S—and, interestingly, such of these as have been subjected to depolymerization by sulphydryl compounds, have lost their activity (Heimer and Frederico, 1958; Grubb and Swahn, 1958). A somewhat similar situation may obtain in rabbits also, since some antibodies against sheep red cells (Stelos, 1956; but *not*, apparently, against human red cells, Johnson, Woernley and Pressman, 1957) are associated with heavy components, particularly at later stages of immunization.

Antibodies may be said to have been found in all the normal gamma globulin components. They have not been found in human pathological globulins, but I have known a rabbit hyperimmunized against pneumococcus type III whose serum contained a cryoglobulin, 40 per cent of which was demonstrably antibody and precipitated with purified capsular polysaccharide. Furthermore, those antibodies whose amino acid composition has been studied chemically, by Porter (1950) and by Emil Smith and his colleagues (Smith, McFadden, Stockell and Buettner-Janusch, 1955), not only had virtually identical amino acid compositions, but also had the same sequence of 5 N-terminal amino acids as had gamma globulin which was not specific antibody. The odds against this being due to chance are very great indeed. It is, I think, fair to state that antibodies have always been shown to be antigenically similar to the normal gamma globulins of the corresponding fraction of serum in which they occur (even though this may be only a small proportion of the whole globulins), and that no real difference has ever been found. Possibly now that Porter (1958) has succeeded in splitting rabbit antibody molecules into three parts—two only of which combine with the antigen, and presumably have the combining groups—some finer differences may be discovered, but so far there are none.

How are all these facts to be put together? The question 'What are gamma globulins?' is still unanswered, and there

prepared by organic solvent fractionation procedures contains many kinds of antibody—this is the rationale for treatment with 'immune globulin'—and this fraction is composed largely of gamma globulin with sedimentation constant 6–7S. However, if the fractions obtained by chromatography of human serum on an anion exchange powder, as shown in Figure 3, are examined for various kinds of antibody, antibodies may be found throughout all the fractions containing gamma globulin—antibodies against certain antigens often appearing characteristically in a given serum in particular positions. Reaginic skin sensitizing antibodies appear in the middle fractions (Humphrey and Porter, 1957)—while some isoagglutinins appear right at the end. When serum of a rabbit hyperimmunized against ovalbumin or against type III pneumococcus is run in the same way a similar sort of pattern emerges, but antibodies against ovalbumin or the capsular polysaccharide are found in all the fractions containing gamma globulin—which seems to imply that there is no fundamental difference between the fractions as regards their capacity to be or not to be antibody against any given antigen. The more carefully work is done on trying to locate and identify certain special antibodies, the more it becomes apparent that antibodies can differ in their physicochemical properties as much as can the components of normal globulins which, though antigenically similar, vary widely in electrophoretic mobility and even in molecular weight. Thus in Man most workers agree that useful antibodies, and such antibodies as 'blocking' antibodies, occur in the slow gamma fraction and have sedimentation constants about 6 3S; so do warm incomplete Rh antibodies (Fudenberg and Kunkel, 1957). On the other hand, reaginic antibodies (Cann and Loveless, 1957; Kuhns, 1954; but see Sehon, Harter and Rose, 1956); anti A and B isoagglutinins (Faure, Fine, Saint-Paul, Eyquem and Grabar, 1955); cold haemagglutinins (Christenson, Dacie and Croucher, 1957); warm complete Rh antibodies (Faure, Fine, Saint-Paul, Eyquem and Grabar, 1955); the Wasserman factor (Davis, Moore, Kabat and Harris, 1945); the rheumatoid arthritis factor which reacts with normal gamma globulin (Franklin, Holman, Muller-Eberhard and Kunkel, 1957), if

the electrophoretic mobility increased, the amounts of hexose, hexosamine, fucose and sialic acid increased also. It is tempting to suppose that the physico-chemical properties of a family of very similar proteins may be partly, at least, controlled by the amount of carbohydrate in them. Since both the protein portion and the carbohydrate portion of the molecule may give rise to distinct antibodies against their various potential antigenic groupings, such an explanation could also account for their antigenic similarities and their differences. This cannot, however, be a complete explanation. Müller-Eberhard and Kunkel subjected whole human gamma globulin to zone electrophoresis in a starch block, and re-ran the different fractions. As shown in Figure 4, they had well-graded but distinct mobilities. Nevertheless, the carbohydrate analyses were almost identical in all but

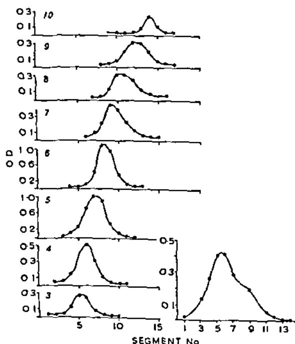


FIG 4 Numbered fractions from whole gamma globulin (lower right-hand corner) were re-run under similar conditions. (Taken from Müller-Eberhard and Kunkel, 1956.)

appears a further question—namely 'Are all gamma globulins antibody?' I suggest that the evidence of immunoelectrophoresis should be accepted that molecules antigenically related to gamma globulins may be found in the beta and even the $alpha_2$-region, and belong to the same family. Some interesting work by Müller-Eberhard and Kunkel (1956) bears on this point. They studied the carbohydrate content of normal and pathological human gamma globulins—the carbohydrate being an important part of the molecule which is often overlooked. Table 3 illustrates their findings with myeloma proteins. As

TABLE 3. Electrophoretic mobilities and carbohydrate contents of human myeloma proteins and whole γ-globulin (Fraction II). (Taken from Müller-Eberhard and Kunkel, 1956)

NO	ELECTROPHORESIS	mg Ninh. / mg Fol	CARB MOLES PER MOLE PROTEIN				HEXOSAMINE / HEXOSE †
			HEXOSE	HEXOSAMINE	FUCOSE	SIALIC ACID	
γ Fr II			10 5±0 4*	10 3±0·3 *	2 0±0 1*	1 0	0 98
XXI		0 58	16 0±0 8*	11 6±0 3 *	2 8	1 9	0 74
XIII		0 55	7 8	8 6	2 0	0 8	1 10
XVI		0 50	6 8	8 7	2 0	0 5	1 28
VI		0 50	10 0	8 6		1 2	0 86
X		0 50	9 0	8 9		0 9	0 99
XVII		0 50	9 0	9 0	2 0	0 6	1 00
XX		0 50	8 3±0 1*	7 5±0 2*	2 0	1 2	0 91
III		0 50	24 2	15 1		1 0	0 62
XII		0 45	25 0	21 2	2 3	3 0	0 85
XV		0 42	26 0	21 1		3 0	0 81
XI		0 42	23 4	18 8		5 7	0 80
XVIII		0 44	22 9	22 4	3 0	4 0	0 98

essentially the same amino acid composition, and the same amounts of hexose and hexosamine, and all were antigenically very similar—although possibly complex. Thus the differences in electrophoretic mobilities must be due to some relatively fine differences in arrangement of the constituent amino acids. This is, of course, not unreasonable. Even a single protein antigen contains multiple potential antigenic sites, and the antibodies against the different sites, and antibodies with different affinities, must differ somewhat in their surface configuration.

If gamma globulins are in fact a family of closely similar molecules with varying surface arrangements of amino acids, and with the possibility of varying amounts of carbohydrate, how does this variation come about? There seem to be three possibilities: that an original standard molecule is produced by a single kind of cell, but that after manufacture it becomes altered (by loss of carbohydrate or by rearrangement) in a host of different ways; that a single kind of cell produces varying kinds of molecules at different times—due to changes in the internal or external environment; or that a family of differing cells (which may all look alike) each produce approximately constant, but differing, kinds of molecules. Askonas, Humphrey and Porter (1956) have produced evidence which makes the first two explanations unlikely, at least so far as concerns gamma globulin synthesis in the rabbit. Our evidence gives some support for the third alternative. It was based on separation of rabbit gamma globulin into multiple overlapping fractions by a technique developed by Porter (1955), which differs essentially from any of the techniques mentioned so far—namely partition chromatography between aqueous phosphate and a mixture of butyl and ethyl cellosolves on a Celite column. The technique is exacting—but, when once working, it gives extremely [illegible] nphrey and Port [illegible] which is simila [illegible] an immunized animal. The figure also shows another phenomenon which we investigated, namely the change in properties of the antibody molecules with continued immunization. The animal whose serum was analysed in this experiment had been

the fastest fraction. Dr. B. A. Askonas and I have been studying what could in some ways be termed abnormal gamma globulins appearing in the plasma of rabbits hyperimmunized with pneumococcus type III. Figure 5 shows the free electrophoretic



FIG. 5 Free electrophoresis of serum from rabbit hyperimmunized with type 3 pneumococci Note the enormous gamma globulin peak.

pattern of the serum—which contains so much gamma globulin as to be more like that from a patient with myeloma. After running for 4 hours the gamma globulin separated into three quite distinct peaks (Fig. 6)—although in normal rabbit serum



FIG. 6. Later stage in electrophoresis illustrated in Fig. 5. Only the gamma globulin remains, and this has split into three distinct peaks

there is only one peak. The three fractions could be separated by electrophoresis in a column of treated cellulose. Each contained about 40 per cent of specific antibody—each had

measured. In any given experiment variations in specific radioactivity of fractions represent variations in the rates of synthesis of those fractions, relative to the overall synthesis of gamma globulin by the particular tissue slice. Quite marked differences

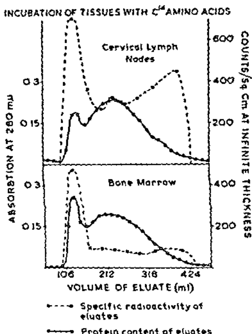


Fig. 8. Specific radioactivity of γ globulin fractions obtained by [illegible]

were found between the behaviour of different tissues. For example a lymph node incorporated labelled amino acids to a greater extent into the fractions of gamma globulin eluted later than into those eluted earlier, whereas the spleen did the opposite (Figs. 8 and 9). A study of incorporation by the spleen of a hyperimmunized rabbit, in which the specific radioactivities of the fractions were measured both before and after specific precipitation of the antibody by antigen, showed that antibody synthesized by the spleen was predominantly in the middle

immunized over a relatively long period of time with ovalbumin and for a short period with pneumococci. The distribution of the two antibodies is quite different, those formed after brief immunization being concentrated in the later fractions, while

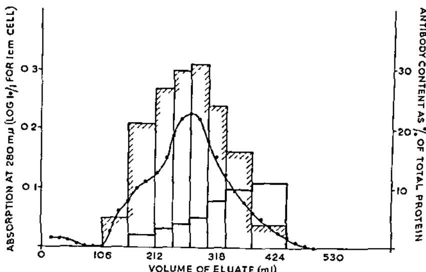


FIG. 7. Gamma globulin was prepared from serum of a rabbit which had received injections of ovalbumin over a long period and of pneumococci over a short period ●——● = protein concentration, hatched columns represent anti-ovalbumin as percentage of total protein, white columns represent anti-pneumococcus antibody as percentage of total protein. (Taken from Humphrey and Porter, 1956)

those formed after prolonged immunization spread into the earlier fractions. A similar change of distribution was observed with antibodies against a bacterial carbohydrate, a protein, and a virus antigen. It is an old observation that the character of antibodies tends to change as immunization proceeds, and it requires an explanation.

The technique of fractionation by partition chromatography was used to study the incorporation of ^{14}C labelled amino acids by slices of various rabbit tissues *in vitro* into different fractions of gamma globulin. Known amounts of whole gamma globulin were added to the slices beforehand, and after incubation the various fractions were isolated and their specific radioactivities

permit a particular kind of cell to proliferate wildly, it could explain the production of myeloma proteins.

The question still remains whether all gamma globulins are antibodies. Any answer is very hard to prove or to disprove,

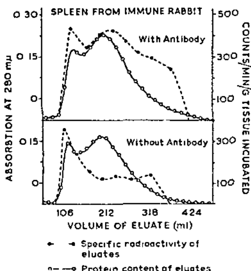
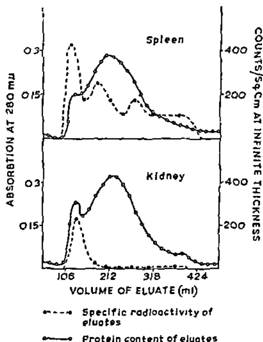


Fig 10 Specific radioactivity of γ-globulin fractions obtained after incubation of spleen slices from an immunized rabbit Above· before removal of antibody, Below: after removal of antibody. The conditions were similar to those in Fig 8.

especially when more and more antibodies against various unsuspected antigens turn up in pooled normal human gamma globulin fractions. Studies on young animals, or on patients with hypoglobulinaemia, certainly indicate that the capacity to produce antibodies and to produce gamma globulin go hand in hand. It might be hoped that a decisive answer would be provided by experiments on germ-free animals, since, if gamma globulins only arise as a result of antigenic stimulation, unstimulated animals might be expected to have none. So far, however, a decisive answer has not been given, for the germ-free animals have produced not only some gamma globulin, but also some antibodies, presumably due to stimulation by dead bacterial products in food or dust (Reyniers, 1957).

fraction of the gamma globulin, while the non-specific gamma globulin synthesized was distributed more evenly (Fig. 10).

These findings are compatible with the idea that different tissues contain different populations of cells making the differ-



ent kinds of gamma globulin. Presumably the relative proportions of the different cells can, and do, change—e.g. when proliferation occurs during prolonged immunization, dependent upon the extent to which the tissue becomes stimulated by the antigen. Any of the cells may be capable of elaborating antibody with a given specificity, though they do not necessarily do so. Such a hypothesis, of course, only pushes the problem a stage further back—but it makes it possible to explain the varying nature of antibodies with continued immunization—and, if we

evidence, that the cells which produced the gamma globulin had migrated into the lung from other parts of the body, the experiment seems to indicate that, in response to an antigenic stimulus, large amounts of new gamma globulin were being formed which did not combine with the antigen. Whether this is termed 'normal' gamma globulin or 'non-specific' gamma globulin is immaterial, but to call it antibody is to stretch this term so far that it becomes meaningless. In our present state of ignorance about antibody formation, I see no reason why plasma cells should not be stimulated to proliferate and to secrete gamma globulin molecules which do not necessarily bear the imprint of any particular antigen.

ACKNOWLEDGEMENTS

I am indebted to the Editors of the *Biochemical Journal*, the *Journal of Clinical Investigation*, Academic Press Inc, and the *Journal of Experimental Medicine* for permission to reproduce Figures 1, 3, 4, 7, 8, 9, 10 and Tables 1, 2 and 3.

REFERENCES

ASKONAS, B. A. and HUMPHREY, J. H. (1958). *Biochem. J.* **70**, 212.
ASKONAS, B. A., HUMPHREY, J. H and PORTER, R. R. (1956) *Biochem. J.* **63**, 412.
CAMPBELL, D H, STURGEON, P. and VINOGRAD, J. R. (1955). *Science*, **122**, 1091.
[illegible] *J. Allergy*, **28**, 379
[illegible] OUCHER, B. E. E. (1957). *Brit. J.* [illegible]
DAVIS, B. D, MOORE, D. H., KABAT, E. A. and HARRIS, A. (1945) *J. Immunol.* **50**, 1.
[illegible]
[illegible] 7, 272.
FARR, R S (1956) *Fed Proc* **15**, 586.
FAURE, R, FINE, J.-M, SAINT-PAUL, M, EYQUEM, A and GRABAR, P. (1955) *Bull Soc Chim biol* Paris, **37**, 783.
FRANKLIN, E C, HOLMAN, H. R., MULLER-EBERHARD, H. T. and KUNKEL, H. G. (1957) *J. exp Med.* **105**, 425 (1957).
FUDENBERG, H H. and KUNKEL, H. G. (1957). *J. exp. Med.* **106**, 689.
GRABAR, P. (1954). *Therapie*, **9**, 163.

Evidence is available to support the idea that not all gamma globulin is antibody—even though it may be synthesized as a consequence of an antigenic stimulus. Like all negative propositions it is difficult to prove, and it depends upon showing that gamma globulin synthesis can be stimulated in such a way that not more than a part of the gamma globulin synthesized is specific antibody. In studies of antibody production by tissues of hyperimmunized rabbits Dr. Askonas and I found that all the tissues which had been stimulated by the antigen to produce antibody also made at least as much non-specific gamma globulin as antibody (Askonas and Humphrey, 1958). To designate any material as non-specific gamma globulin is, of course, merely to state that we were unable to cause the material to precipitate or to co-precipitate with the antigen or to find evidence that it was capable of binding with the antigen—e.g. by Farr's method (Farr, 1956) of precipitating the globulin with ammonium sulphate in the presence of radioactive antigen. It could be argued that the non-precipitable gamma globulin consisted of antibodies to unknown antigens and that its production was stimulated by some form of anamnestic reaction. However, when we studied the *in vitro* synthesis of antibody against type III pneumococci in the isolated perfused lung of an immunized animal we again observed synthesis of at least as much non-specific gamma globulin as of antibody against type III capsular polysaccharide. This antibody normally comprises at least 90 per cent of the antibody directed against any of the components of sonically disintegrated type III pneumococci, when rabbits are immunized with formalin-killed organisms by the technique which we used. Furthermore, after removal of the anti-capsular polysaccharide antibodies, the remaining gamma globulin in the perfusate contained only minimal amounts of precipitable antibody against other components of the disintegrated pneumococci. The non-specific gamma globulin, therefore, was not likely to be unrecognized antibody against the antigen which stimulated its production. Nor was it antibody which would normally have been made in the lung, since lungs taken from normal animals synthesize only small amounts of gamma globulin. Apart from the possibility, for which there is no

VII

Bacterial Genetics and Gene Structure

W. HAYES

THE aim of this paper is to reveal something about the methods and potentialities of microbial genetics and about the contributions which this very new branch of genetics has made, and is making, to knowledge of biological organization at the cellular level. Within recent years we have witnessed many important incursions into the exciting field of molecular biology which seeks to explain living processes in the ultimate terms of the behaviour of atoms and molecules. The earliest and most striking successes of this fundamental approach are seen in the detailed analyses of the chemical steps involved in biosynthesis and respiration. But such activities of the cell comprise only a part of the phenomenon of life. To consider them in isolation is like studying the various agricultural and industrial pursuits of human society without reference to the system of government and administration which gives these activities cohesion and continuity. The study of the government of the cell is, of course, genetics, but it is only within the last few years that a significant break-through to the molecular level has occurred in the study of genetic phenomena. The initial, and perhaps most important, advance was the identification of the chemical nature of the gene with deoxyribonucleic acid or DNA. This identification is supported by two kinds of evidence, indirect and direct. The indirect evidence is, briefly, as follows.

(1) DNA is found in considerable amounts in all cells, whether animal, plant or bacterial, and in these cells it is virtually restricted to the nucleus which is the site of the genetic apparatus of the cell.

[illegible] 205.
[illegible] 79, 234.
[illegible] Med. 106, 467.
[illegible] 231, 1119
[illegible] exp. Med. 104, 253
[illegible] exp. Med. 108, 121.
[illegible] Soc. 78, 751.
[illegible]

PUTNAM, F. W. (1957). *Physiol. Rev.* 37, 512.
REYNIERS, J. A. (1957). *Amer. J. vet. Res.* 18, 678.
SEHON, A. H., HARTER, J. G. and ROSE, B. (1956). *J. exp. Med.* 103, 679.
SLATER, R. J., WARD, S M. and KUNKEL, H. G. (1955). *J. exp. Med.* 101, 84.
SMITH, E L , MCFADDEN, M. L., STOCKELL, A. and BUETTNER-JANUSCH, V. (1955). *J. biol Chem.* 214, 197.
SMITHIES, O (1955) *Biochem. J.* 61, 629.
STELOS, P. (1956). *J. Immunol.* 77, 396
TREFFERS, H. P (1944). *Adv in Protein Chemistry*, 1, 88.
WILLIAMS, C. A. and GRABAR, P. (1955). *J. Immunol.* 74, 404.
WOODS, K. R., PAULSEN, E. C , ENGLE, Jr., R L. and PERT, J. H. (1958). *Science*, 127, 519
WURMSER, R. and FILITTI-WURSMER, S. (1957). *Progress in Biophysics*, 7, 87.

transformation, is found in *Pneumococcus*, *Haemophilus* and several other genera, and involves such diverse characters as specific polysaccharide synthesis, the capacity to ferment various carbohydrates, and resistance to a wide range of antibiotics and chemotherapeutic agents. Molecular weight determinations of transforming DNA preparations show that the size of the transforming particles lies between about one-thousandth and one ten-thousandth that of the bacterial nucleus. From this it follows that if the donor and recipient strains differ in two characters it is highly unlikely that the genes determining these two characters will lie so closely together on the chromosome as to be carried by the same DNA particle. Experimentally, while some cells of the recipient population may be transformed with respect to one character, and others with respect to the other character, it is rare for any one cell to be transformed with respect to both characters at the same time. There is good evidence that a molecule of transforming DNA is not just added to the genetic machinery, but actually replaces a counterpart on the chromosome of the recipient cell. For example, it has been shown that reciprocal transformation can occur between two strains which differ from one another in some observable way as a result of a single mutation; thus not only can DNA from type 1 cells transform type 2 cells to type 1, but DNA from type 2 cells can also transform type 1 cells to type 2. Moreover, DNA extracts of the progeny of the cells so transformed carry the transformed character but not the alternative character which the ancestors of these cells possessed before transformation. The simplest explanation is that the incoming DNA molecule is not just a dominant gene which is added to the recipient cell, but that it actually replaces its homologue on the recipient chromosome. Indeed there are cases of what are termed 'allogenic transformation' in which a single transformation event results in the segregation of different recombinant progeny (Ephrussi-Taylor, 1951). These cases happen to involve rather complex characters which I cannot detail here. They are explicable, however, in a quite orthodox way, by supposing that the character concerned is controlled by at least two genes, which are located so closely together on the bacterial chromosome

(2) All the somatic cells of any given species contain a very constant amount of DNA irrespective of their functional differentiation or metabolic state. The amount of ribosenucleic acid (or RNA), on the other hand, varies markedly with metabolic activity.

(3) The amount of DNA present in haploid germ cells, which contain only one set of chromosomes, is just half that found in the diploid somatic cells which possess two sets of chromosomes. Similarly, when the ploidy, or number of chromosomes, is still further increased, the amount of DNA per cell rises proportionately.

This kind of evidence, however, while strongly suggesting that DNA is principally associated with the chromosomes, is far from providing proof that it actually carries the genetic information. DNA is usually found in the nucleus in the form of nucleoprotein and it seemed to some to be more probable that it was the *protein* moiety of the nucleoprotein that conferred genetic specificity in view of the great number of unique configurations that protein can adopt.

DNA as the Chemical Basis of Heredity

(1) *Transformation*

The direct evidence for DNA was much more convincing and came from two different kinds of experiment in microbial genetics. The first was the discovery (Griffith, 1928; Avery, MacLeod and McCarty, 1944) that virtually pure, highly poly-

tracted from streptomycin-resistant donor cells is mixed with streptomycin-sensitive recipient cells, a proportion of these initially sensitive recipient cells is found to have become resistant, and to transmit the character of resistance to all their *progeny*. The ability of the DNA to transmit characters in this way is completely and specifically destroyed by treatment with the enzyme deoxyribonuclease. This phenomenon, known as

the DNA. Hershey and Chase (1952), in their classical experiment, specifically labelled the protein coat of the virus with radioactive sulphur, ^{35}S, and its DNA with radioactive phosphorus, ^{32}P. They showed that, on infection of sensitive bacteria

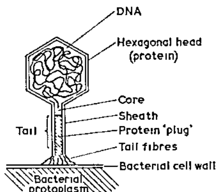


FIG 2 Diagrammatic representation of the structure of a bacterial virus (bacteriophage) particle.

with such particles, only the virus DNA and about 3 per cent of its protein enters the bacterial cell. The protein sheath remains outside the cell and, once the DNA has been injected, can be stripped off with a blendor without altering the course of the infection. Once injection has occurred, the DNA behaves like a 'master gene' which turns over the biochemical machinery of the cell to the exclusive manufacture, not only of virus DNA, but also of new virus protein, so that, about 20 minutes later, the cell bursts and liberates several hundred mature virus particles.

It is a pity that the 3 per cent of injected protein spoils the perfection of this evidence that DNA is the only possible repository of genetic information in bacterial viruses. There are, however, two other lines of supporting evidence. The first is that the infectivity of phage particles is sensitive to the decay of ^{32}P atoms incorporated exclusively into their DNA. For every transmutation of an incorporated ^{32}P atom into ^{32}S there is about a

that they are often transmitted together on the same particle of DNA. In Figure 1 the two genes are called A and B; the character controlled by these genes is altered, in the recipient strain by mutation of gene A to a, and in the donor strain by

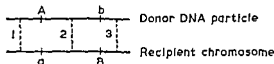


FIG. 1

mutation of gene B to b. If we assume that a transforming DNA particle is actually a minute fragment of donor chromosome which pairs with its homologue in the recipient cell, then genetically different progeny will result depending on the region of donor chromosome involved in the recombination process. Experimentally, recombinant progeny of type AB (resulting from crossing-over in positions 1 and 2) and Ab (resulting from crossing-over in positions 1 and 3) were found.

We are therefore justified in regarding transformation as a sexual process which differs from sexual processes in more highly evolved organisms mainly in the somewhat artificial method of insemination, and the fractional genetic contribution of one of the parents. As we shall see, this inequality in the parental contributions to the zygote is characteristic of all types of bacterial sexuality.

(2) *Bacteriophages*

Let us now turn from bacteria to quite a different kind of organism, if you can call it an organism, namely, virulent bacterial viruses or bacteriophages. So far as chemical analysis and electron microscopy can determine, phages are very simple creatures (Fig. 2). They appear to consist of a protein sheath in the form of an hexagonal head and tail, splayed out into fibrils at its tip, whereby the virus attaches itself to the wall of the bacterial cell. The head encloses pure DNA in the form of a long filament. The virus particle is devoid of RNA. Within the lumen of the tail is a protein plug which one can imagine as corking in

the DNA. Hershey and Chase (1952), in their classical experiment, specifically labelled the protein coat of the virus with radioactive sulphur, ^{35}S, and its DNA with radioactive phosphorus, ^{32}P. They showed that, on infection of sensitive bacteria

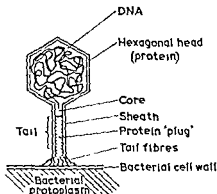


FIG. 2 Diagrammatic representation of the structure of a bacterial virus (bacteriophage) particle.

with such particles, only the virus DNA and about 3 per cent of its protein enters the bacterial cell. The protein sheath remains outside the cell and, once the DNA has been injected, can be stripped off with a blendor without altering the course of the infection. Once injection has occurred, the DNA behaves like a 'master gene' which turns over the biochemical machinery of the cell to the exclusive manufacture, not only of virus DNA, but also of new virus protein, so that, about 20 minutes later, the cell bursts and liberates several hundred mature virus particles.

It is a pity that the 3 per cent of injected protein spoils the perfection of this evidence that DNA is the only possible repository of genetic information in bacterial viruses. There are, however, two other lines of supporting evidence. The first is that the infectivity of phage particles is sensitive to the decay of ^{32}P atoms incorporated exclusively into their DNA. For every transmutation of an incorporated ^{32}P atom into ^{32}S there is about a one-in-ten probability that the phage is killed, in the sense that it can no longer direct the synthesis of new phage DNA or protein. The second line of evidence comes from the study of what

is called *phenotypic mixing*. If a bacterium is simultaneously infected with two phage particles, A and B, which differ in the specificity of their tail proteins, the infected cell synthesizes both 'A' and 'B' protein as well as 'A' and 'B' DNA. However, when the time comes, towards the end of the latent period, for the protein and the DNA to be assembled into mature phage particles, it seems that the protein envelopes cannot readily distinguish between homologous and heterologous DNA. The result is that although the majority of emerging particles are normal types A and B, in an appreciable proportion type 'A' DNA is enclosed in a type 'B' protein coat and vice versa. When, however, susceptible bacteria are infected with such phenotypically mixed particles, for example with particles having 'B' DNA enclosed in 'A' protein coats, the particles which emerge from this second infection are of pure type 'B'; that is, the constitution of the progeny is determined solely by the DNA, and not by the protein of the infecting parental particle.

. . . .

are composed of protein and RNA and are devoid of DNA. Tobacco mosaic virus, for example, consists of a protein rod with an axial hole running through it, the RNA being embedded in the protein periaxially. It was found by Fraenkel-Conrat, Singer and Williams (1957) that when the virus is fractionated chemically, so that its protein and RNA are separated from one another, the RNA is infective alone (though to a much lower level than its equivalent in the original virus) while the protein is not infective. However, if the purified protein and RNA are mixed together in a test-tube, fully infective virus particles reform automatically. Various strains of tobacco mosaic virus exist which differ from one another mainly in the antigenic specificity of their protein components. If two strains of virus, A and B, are chemically fractionated and the protein of one strain mixed with the nucleic acid of the other, virus rods are reconstituted as before. These have the structure shown in Figure 3.

When the plant is infected with these viruses, the progeny

viruses which are liberated are found to have protein of the same type as the nucleic acid, and of different type from the protein, of the infecting particles. Thus the progeny of Virus 1 have protein 'B' and RNA 'B', and those of Virus 2, protein 'A' and RNA 'A'.

Virus 1 Protein 'A'+RNA 'B'
Virus 2 Protein 'B'+RNA 'A'

FIG. 3

I think two conclusions may now be drawn from what I have said. The first is that there is now very good circumstantial evidence that the genes, of bacteria and their viruses at least, are composed exclusively of DNA. The second is that there is, I think, almost conclusive direct evidence that both DNA and RNA can carry all the information necessary for their own as well as for specific protein synthesis.

SYNTHETIC GENETICS

(1) DNA *Structure and replication*

Having got this far, the next logical step is to consider DNA at the structural level, and try to find an answer to two vital questions which lie at the very foundation of biology—how does it carry the genetic information, and how does it reproduce itself? This brings me to the famous Watson-Crick model of DNA structure (Watson and Crick, 1953). From the purely chemical point of view, DNA is built up of a chain of alternating deoxyribose sugar and phosphate groups as shown in Figure 4a, to each sugar is attached a nitrogenous base of which four kinds are usually found—the two purines adenine (A) and guanine (G) and the two pyrimidines thymine (T) and cytosine (C). Each unit consisting of base-sugar-phosphate is called a nucleotide unit and the chain is called a polynucleotide chain. The Watson-Crick model, constructed from chemical and X-ray diffraction data, comprises two polynucleotide chains, joined together by hydrogen bondings between the purine and pyrimidine bases which face inwards towards each other, while the sugar-phosphate chains form backbones on the outside; this can

be represented as a ladder (Fig. 4b) of which the sugar-phosphate chains form the uprights while the pairs of bases form the rungs. Finally, the two chains are wound helically around one another to form a braid, as if you took the two ends of the ladder and twisted them in opposite directions (Fig. 5).

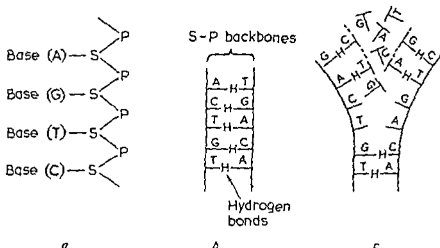


FIG. 4 Diagrammatic representation of the physico-chemical structure of deoxyribonucleic acid (DNA)

It is essential for the regularity of this structure that the bases of the two chains be paired off in a highly specific way, such that the adenine of one chain is always joined to the thymine of the other, and the guanine to the cytosine, as shown in Figure 4b. So far as any one chain is concerned, however, there is no restriction whatever on the *order* in which the bases are arranged longitudinally, although it will be obvious from the specificity of pairing that the order on one chain determines that on the other.

From the biological point of view this structure has two very appealing features. Firstly, it at once suggests a plausible code for the carriage of genetic information. Since the backbone is highly regular, this code must reside in the arrangement of the bases, and the most obvious straightforward code would be the sequential order of A, G, T and C along one of the polynucleotide chains. Crick and his colleagues (Crick, Griffith and Orgel,

1957) have recently shown, as a problem in pure coding, that a code of non-overlapping triplets of nucleotides could carry the information required for the synthesis of twenty amino acids, and could therefore determine the structure of any biological protein we may care to specify.

The second important feature is that this structure appears well adapted to self-replication since the arrangement of bases on the two strands is complementary, so that each can act as a template for the assembly of the opposite strand.

(2) *Experimental approaches to the mechanism of replication*

The next question is 'Is it possible to investigate experimentally how replication actually occurs?' The most obvious experimental approach would be to label the DNA of living bacteria, or virus particles, and then to observe how the labelled atoms are distributed among the DNA molecules of the progeny after replication has occurred. Three main theoretical mechanisms for replication have been proposed which should give distinguishable results in labelling experiments. According to Watson and Crick, all that is required is that the hydrogen bonds joining the complementary bases should break and that the two strands should then unwind and separate, as shown in Figure 4c. Newly synthesized, free nucleotide units could then attach themselves specifically, by hydrogen bonding, to their complementary bases and be polymerized to form two replicas of the original double helix. The essence of this scheme is that each strand of the original double helix acts as a template for the assembly of a newly synthesized strand. Thus, if the parental double helix was labelled, at the end of the first generation each daughter duplex should have one labelled and one unlabelled strand, that is, exactly one half the amount of labelling of the parental duplex. Of the four granddaughter duplexes at the end of the second generation, two should have one labelled and one unlabelled strand as before, while the other two should have both strands unlabelled (Fig. 6, scheme 1).

Unfortunately at this point we run into a very serious snag. It is that the two coils are wound around one another plectonemically, that is, that they are interwoven like a braid, and can

only be separated in one of two main ways. The first is by rotation round their axes, that is, by unwinding. The amount of unwinding required would be one complete turn for every ten nucleotide units. Since the DNA of something as small as one of

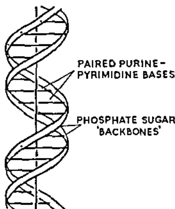
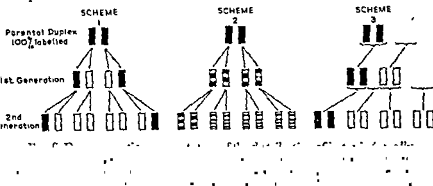


FIG. 5. Diagrammatic representation of the double helical arrangement of the two polynucleotide chains of DNA. (After Watson and Crick, 1953.)

the 'T' series of coli-phages contains several hundred thousand nucleotide units and is replicated in a matter of a minute or so, this method of unravelling seems inherently improbable—all the more so when we consider that the DNA fibre is very long and must be folded in a complicated manner in order to accommodate it within the confines of the infected bacterium in which it replicates. However, Levinthal and Crane (see Delbruck and [illegible] d for [illegible] of a [illegible] ly a fraction of that required for polymerization of the polynucleotide backbone, so we must regard this method of replication as definitely 'on the map'.

An alternative method of unravelling the duplex, which gets over the unwinding difficulty, was proposed by Delbruck (1955). In this scheme (Fig. 6, scheme 2) the two strands unravel by repeatedly breaking during replication, and then reuniting with the appropriate ends of newly synthesized strands. This

means that, on replication, each strand will give rise to two daughter strands, in each of which segments of labelled and unlabelled DNA will alternate. As scheme 2 shows, there will be no unlabelled strands in either the daughter or granddaughter duplexes.



(1953); scheme 2 is that of Delbrück (1955), scheme 3 involves transfer of information from DNA to a (?) protein template, without separation of the two strands of the DNA double helix. Each of the paired rectangular blocks represents one strand of the DNA double helix. The hatching represents the distribution of marker atoms among the strands of daughter double helices.

The third scheme (Fig. 6) suggests a way in which replication could occur without separation of the two strands of the duplex. If you looked at a model of the double helix of DNA you would see that around the outside of the model, between the backbones, run two helical grooves. Evidence by Wilkins and his colleagues (Feughelman *et al.*, 1955) suggests that, in nucleoprotein, protamine is accommodated in the smaller of these grooves. It is possible that the DNA duplex as a whole could transfer its information to either protein or RNA lying in these grooves, which, in turn, could act as a template for the synthesis of a complete new duplex. In this third scheme (Fig. 6), the original duplex would remain intact, and not transfer any of its labelling so that, at the end of the first and second generations there will be one duplex with both of its strands fully labelled while the other duplexes are unlabelled.

Attempts, especially by Stent and Levinthal, to follow the distribution of radioactive ^{32}P atoms in the DNA of virulent

phages during their replication in infected bacteria, have yielded somewhat inconclusive results. Probably the most important complicating factor in these experiments is that the DNA molecules of these phages mate as they replicate in the DNA pool. The matings appear to be random, and are accompanied by genetic reassortments, that is, by recombination, which probably leads to a redistribution of the ^{32}P label, thus blurring or obliterating the replication picture (see Delbruck and Stent, 1957).

The first experiment to give an apparently clear-cut result is that of Meselson and Stahl (1958), who employed a novel and ingenious technique. If a solution of caesium chloride is centri- [illegible]

caesium which tends to disperse it. This leads to a density gradient in the solution. It so happens that the density of DNA molecules lies within this gradient so that if DNA is mixed with the caesium and centrifuged, it will float at its own density level, like a hydrogen balloon in air, and will thus be concentrated into a discrete band. The method is of such sensitivity that if a mixture of two DNA preparations, one containing ordinary nitrogen, ^{14}N, and the other heavy nitrogen, ^{15}N, is analysed, the two DNA fractions will separate out into clearly defined bands, just as balloons filled with hydrogen and helium respectively will float at different levels in air. The experiment is as follows. Bacteria are grown for many generations in a medium containing ^{15}N as the only nitrogen source, so that *all* the ^{14}N of the cells, including that in the DNA, becomes replaced by ^{15}N. At the start of the experiment the organisms are transferred to a fresh medium containing only ^{14}N, and are allowed to grow in this medium for two generations. At intervals after transfer samples are removed, the DNA of the cells is extracted, and the ^{14}N and ^{15}N content of the DNA molecules is analysed. *At the time of transfer to the ^{14}N medium, that is at the start of the experiment, all the DNA molecules are of ^{15}N type* (Fig. 7). At the end of the first generation, all the ^{15}N molecules have disappeared and have been replaced by twice the number of

'hybrid' molecules which contain equal amounts of ^{15}N and ^{14}N, and form a band intermediate in position between pure ^{15}N and pure ^{14}N molecules. At the end of the second generation the number of hybrid molecules remains the same, but a number of unlabelled, pure ^{14}N molecules have appeared and thereafter increase in number. Comparison of Figure 7 with the

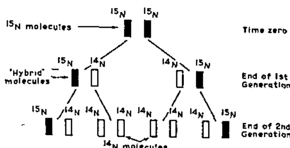


FIG. 7. Distribution of ^{15}N marker atoms among daughter 'molecules' of replicating DNA, as revealed by the experiment of Meselson and Stahl (1958) (see text). This figure should be compared with the three schemes for DNA replication shown in Fig. 6.

three theoretical schemes for replication (Fig. 6) shows that this experimental result conforms only to the Watson-Crick scheme, and is at variance with both the others. If we make the assumption, and this has not been proved, that the DNA molecules defined by this experiment are, in fact, Watson-Crick double helices, then it is difficult to escape the conclusion that, in this case at least, DNA replicates as postulated in Scheme 1 (Fig. 6). At present it seems unlikely that any simple generalization of the problem of replication will be found. For example, there is highly suggestive, but as yet inconclusive, evidence from ^{32}P experiments with phage that during the synthesis of new phage DNA by infected cells, there is an intermediate stage when the genetic information coded in the parental, infecting DNA is transferred to something which is *not* DNA and may be protein (Stent, 1955).

Analytical Genetics—Analysis of Fine Structure

(1) *Theoretical considerations*

lished science of genetics, however, is mainly *analytical* in that it has had to take as its starting-point the crudely observable ways in which characters are inherited among whole organisms and then, by increasing renfinements of technique, to try, inductively, to define the nature of the ultimate determinants of these characters, their organization and their interactions. These two approaches, the synthetic and the analytical, are clearly complementary, and there must be a point at which they will meet at the same level, so that their findings may be equated. I want to show now that we have almost reached this point. But first of all it is necessary to explain some fundamental theory about analysis of the fine structure of the chromosome and gene.

In order to study the order of arrangement of genes on a chromosome, and the relative distances between them, it is necessary to have two parental strains which differ from one another in the characters controlled by these genes. Prior to genetic exchange, or recombination, the two parental chromosomes pair in such a way that genetic loci controlling the same functions, that is, allelic loci, are in apposition. The frequency with which recombination occurs between two genetic loci, resulting in a reassortment among the progeny of the characters controlled by them, is roughly proportional to the distance between them on the chromosomes of the two parents. Thus, if the chromosome of one parent carries the mutation 'a' at the 'A' locus (Fig. 8) and the chromosome of the other parent carries the mutation 'b' at the 'B' locus, and if we assume that every point on the paired chromosomes carries an equal chance of exchange or crossing-over, it is clear that when A and B are widely separated (as in Fig. 8a) there is a high probability that a cross-over will occur on the region of chromosome separating them, so that recombinants 'AB' and 'ab' will frequently be

found. But if the loci are close together (as in Fig. 8b) the probability of a cross-over between them will be proportionately low, and the occurrence of recombinants for these characters correspondingly rare.

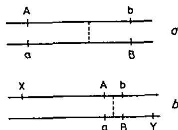


FIG. 8

Theoretically, the dimensions of a genetic locus can be defined as the longest piece of chromosome which cannot be subdivided further by recombination, but such a definition is obviously an arbitrary one. For example, we can pick out, from the progeny of a cross, recombinants for the two rather distant loci, X and Y (Fig. 8b), and examine them to see how often they are also recombinants for the closely linked loci A and B, that is, are of genotype XABY. In other words, we see how often a crossover between X and Y occurs in the short region AB. This gives the frequency of recombination in the region AB as a proportion of the frequency for the much longer region XY. If we find one XABY recombinant for every 100 XY recombinants, then the distance between A and B is about 1/100 that between X and Y. But if we examine 10,000 XY recombinants and fail to find any recombinations for A and B among them, we cannot say that the points A and B coincide, but only that the distance between them is less than one ten-thousandth that of the distance XY. If we examined a million XY recombinants we might find a few XABY progeny among them, and thus set a new upper limit to the length of the region AB.

Now the chromosome of a phage particle consists of several hundred thousand nucleotide units. The chromosome of the bacterium *Escherichia coli* contains about 100 times as much DNA and is, therefore, of the order of ten million double nucleotides

long. It follows that if we wish to extend genetic analysis to the molecular level in bacteria in order, for example, to find out how many nucleotide units are genetically meaningful, two requirements must be met:

(*a*) We must have mutations on the parental chromosome which are of the order of only one ten-millionth the length of the chromosome apart. Otherwise there is nothing on the chromosome to measure. But since mutations are rather random events with regard to their distribution along the chromosome, this means that very large numbers of mutants must be isolated, and the position of the mutations mapped before we are likely to find some situated so closely together. This problem can be greatly simplified, of course, by working with phage, whose chromosome is 100 times shorter, or as we shall see, by selecting for mutations involving a character which we know to be controlled by only a very small chromosomal region.

(*b*) We must also be able to isolate and identify the exceedingly rare recombinants which arise from crossing-over between these mutations, that is, with a probability of the order of one in ten million in the case of bacteria or of one in 100,000 in the case of phage.

It is only in micro-organisms, and especially in bacteria, that such exacting experimental requirements can readily be met. This is because, in them, parental strains can be marked by precise differences in chemical function, enormous populations can rapidly be mated under controlled conditions, and, from such populations, rare recombinant types can easily be isolated by selective methods, instead of by inspection. For example, in bacteria and fungi the mutations most widely used in the analysis of fine structure are those involving loss of some simple synthetic function such as the ability to synthesize tryptophan. If a pair of tryptophan-requiring mutant strains, having their mutations at different sites on the chromosome, are mated on synthetic agar devoid of tryptophan, the only cells which can multiply and produce colonies will be those progeny of the cross in which the ability to synthesize tryptophan has been reconstituted by recombination.

(2) *The bacterial chromosome—transduction*

The most extensive and important work on the fine structure of the bacterial chromosome has been done by Demerec and his colleagues of the Carnegie Institute of Washington, using *Salmonella typhimurium* as the organism and transduction as the tool. In transduction, a bacterial virus of low virulence acts as a genetic vector, transporting a small fragment (about one-hundredth part) of the chromosome of the donor cell, which liberates the virus, into a recipient cell which the virus subsequently infects (Zinder and Lederberg, 1952). In the recipient cell the fragment of donor chromosome pairs with the corresponding region of the recipient's chromosome with the result that a recombinant cell may be formed.

The mode of investigation was to isolate, by means of a semi-selective technique, a large number of mutant strains having a requirement for the same metabolic end-product, such as the amino acid tryptophan, for example, or histidine. The relative positions of the sites of all these mutations on the chromosome were then mapped by observing the frequencies with which recombinants not requiring tryptophan arose when the mutants were crossed with one another in various combinations. The kind of findings that have been obtained can best be illustrated by the results of Hartman's work on a series of 44 histidine-requiring mutants (Hartman, 1957). Each number in Figure 9 represents an independently isolated mutant and the diagram shows the relative positions of their sites of mutation on the small fragment of chromosome on which they lie. The four pairs of mutants identified by a black circle failed to give recombination with one another and so could not be allotted a different site; but in the case of three of these four pairs the mutants could be distinguished from each other by other criteria, such as reverse mutation rate. This leaves only two mutants out of the 44 which *might* be identical. This means that the number of possible distinct mutations involving the same general functional defect must be very large indeed. The first point to be made, then, is that the small region of chromosome controlling histidine synthesis can be subdivided into a very large number of distinct mutational *sites*, all of which are closely linked to one another.

long. It follows that if we wish to extend genetic analysis to the molecular level in bacteria in order, *for example*, to find out how many nucleotide units are genetically meaningful, two requirements must be met:

(*a*) We must have mutations on the parental chromosome which are of the order of only one ten-millionth the length of the chromosome apart. Otherwise there is nothing on the chromosome to measure. But since mutations are rather random events with regard to their distribution along the chromosome, this means that very large numbers of mutants must be isolated, and the position of the mutations mapped before we are likely to find some situated so closely together. This problem can be greatly simplified, of course, by working with phage, whose chromosome is 100 times shorter, or as we shall see, by selecting for mutations involving a character which we know to be controlled by only a very small chromosomal region.

(*b*) We must also be able to isolate and identify the exceedingly rare recombinants which arise from crossing-over between these mutations, that is, with a probability of the order of one in ten million in the case of bacteria or of one in 100,000 in the case of phage.

It is only in micro-organisms, and especially in bacteria, that such exacting experimental requirements can readily be met. This is because, in them, parental strains can be marked by precise differences in chemical function, enormous populations can rapidly be mated under controlled conditions, and, from such populations, rare recombinant types can easily be isolated by selective methods, instead of by inspection. For example, in bacteria and fungi the mutations most widely used in the analysis of fine structure are those involving loss of some simple synthetic function such as the ability to synthesize tryptophan. If a pair of tryptophan-requiring mutant strains, having their mutations at different sites on the chromosome, are mated on synthetic agar devoid of tryptophan, the only cells which can multiply and produce colonies will be those progeny of the cross in which the ability to synthesize tryptophan has been reconstituted by recombination.

from the point of view of aminoacid synthesis, appear to be functionally identical. If we equate the gene with the site, this appears to require a modification of the 'one gene, one enzyme' hypothesis, in that a large number of mutational sites are all concerned with the production of a single enzyme. We must remember, however, that enzymes are complex proteins which must themselves be manufactured before aminoacid synthesis can start. The factory and its jigs must be constructed before it can begin to produce. It therefore seems probable that more refined analysis will reveal a *sequential order of sites* determining the sequence of aminoacids in the specific enzyme produced by the functional locus, this sequence being ultimately laid down by the order of nucleotides in the DNA chain. In this way, fine structure analysis offers an experimental approach to decoding the DNA code, by enabling alterations in the aminoacid sequence of an enzyme to be correlated with alterations in the position of mutational sites on the DNA chain.

This investigation of the histidine loci is the most refined and arduous that has yet been attempted in bacteria. To carry it significantly further would require the independent isolation, characterization and mapping of some 500 to 1,000 histidineless mutants which would be a Herculean task. In the event about 50 sites have been mapped within a region of chromosome about 100,000 nucleotides long. There is evidence that these 50 sites are probably not distributed over the entire length of the donor fragment carried over by the phage vector but are compacted into a part of it, let us say one-tenth or 10,000 nucleotides long. This gives us a degree of resolution of about 200 nucleotides between each mutational site.

(3) *The bacteriophage chromosome*

I now want to leave bacteria for a short space to discuss a quite different kind of system studied by Benzer (1955, 1957), which, even allowing for a 100-fold error in estimation, is still potentially capable of distinguishing two points less than one nucleotide apart. Bacteriophage T4 can infect and lyse two strains of *Escherichia coli* called B and K. When the phage is grown on B it produces, at quite high frequency, mutants known as 'r_{II}'

The second point is that when the mutants were characterized in terms of the specific steps in the synthesis of histidine which they were unable to perform, they were found to fall into seven functional groups marked A–G in Figure 9. All the mutational sites within any one group, having precisely the same functional defect, were linked together to form a *locus*; for example, all the seven sites concerned with the conversion of imidizole-acetol phosphate ester to L-histidinol phosphate ester lie together within locus C.

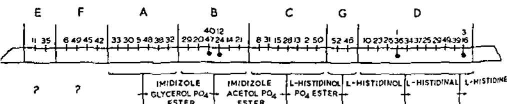


FIG. 9. *Diagrammatic 'map' of the small region of chromosome of Salm. typhimurium* controlling the synthesis of histidine. The numbers indicate the sites of independently isolated mutations inhibiting histidine synthesis. The black circles represent the sites of two mutations not separable by recombination Each of the subdivisions of the map indicated by the letters A–G delimit those mutational sites which block the same step in synthesis (i.e. control synthesis of the same enzyme) (After Hartman, 1957.)

Thirdly, it was found that the various functional loci were linked together in the same order as that of the biochemical steps of histidine synthesis which they controlled. This sequential or 'production line' order has also been found for tryptophan synthesis in both *Salmonella typhimurium* and *Escherichia coli*. On the other hand, in the case of methionine and cysteine synthesis, the order of the loci and of the synthetic steps is not the same. One can only theorize as to the meaning of the production line order of functional loci in certain cases; at present it seems unlikely that it has anything more than an evolutionary significance.

To summarize, then, we find that a small *region* of chromosome concerned with the synthesis of a single aminoacid can be subdivided into a number of functionally distinct *loci*, each responsible for a single synthetic step, and that these loci can themselves be divided into a large number of mutational *sites* which,

light, as a by-product of this classical study, is that the distribution of mutations along the r_{II} region is far from uniform. Mutations tend to pile up at certain sites which Benzer calls 'hot spots', and are very rare at others, suggesting that certain parts of the nucleotide chain, perhaps specific nucleotide pairs, are inherently less stable than others. More recently, Benzer and Freese (1958) have mapped a series of mutations in the r_{II} region obtained by growing phage T4 in cells fed with 5-bromouracil, which is an analogue of thymine and can specifically replace this pyridine base in DNA. As before, the distribution of these bromouracil-induced mutations was far from random; in addition, however, their sites showed no coincidence with those of the first series of spontaneous mutations. On the other hand, a third series of mutations produced by proflavine (Brenner, Benzer and Barnett, 1958), which possibly acts on the phosphate groupings of the DNA backbone, approximated to a Poisson distribution although the sites involved were again mainly different from those of the other two series. The meaning of these interesting findings is not yet known but it is clear that any theory of coding or mutagenesis must take them into account.

Bacterial Conjugation—Measuring the Chromosome in Absolute Terms

I would like to conclude by mentioning a quite different and novel kind of experiment, but one which is very relevant to my theme since it directly relates the genetic organization of the chromosome as a whole to its physical structure. The most highly developed form of sexuality in bacteria is the conjugation mechanism found in *Escherichia coli*. In conjugation, a physiologically differentiated donor cell fuses with a recipient cell and then transfers to the recipient cell a considerable part of its chromosome (Wollman, Jacob and Hayes, 1956). Rarely, the whole of the donor chromosome appears to be transferred to form a complete zygote. Wollman and Jacob (1955, 1958) studied the kinetics of chromosome transfer by separating the mating pairs, at intervals after mixing, by means of violent agitation in a blendor, and then examining the recombinants for inheritance of various markers from the donor parent. They

which can be clearly identified by the altered appearance of their plaques and which are characterized by inability to grow and form plaques on K (Fig. 10). If strain B is mixedly infected with two non-identical r_{II} mutants, wild-type, r^+, particles

Strain of phage T4	Ability to produce plaques on *E. coli* strain: B	K
Wild-type (r^+)	+	+
Mutant r_{II}	+	−

FIG. 10. Shows the ability of phage T4 and r_{II} mutants derived from it to multiply in and lyse the cells of two strains of *E. coli*, B and K (see text).

can be formed by recombination and these can be identified and counted among the progeny of the mixed infection, since they are the only particles which can grow and form plaques when plated on K; the total number of particles is, of course, found by plating high dilutions on B. The sensitivity of the method is such that, using relatively stable mutants, one recombinant among 100 million progeny can be scored. Moreover, this system offers a simple test of the functional similarity or dissimilarity of two r_{II} mutations. If K cells are infected with either mutant alone, or are mixedly infected with two mutants which have the same functional defect, no multiplication of phage occurs: but if the two mutants have different but complementary defects, then they can, as it were, cross-feed one another, multiply and lyse the K cells. Benzer has so far isolated over 900 independent r_{II} mutants, of which a large number have been genetically mapped and their relationships studied. These mutants fall into two functionally distinct groups, the mutational sites within each group being linked together on adjacent segments of the small r_{II} region of the single phage chromosome, just as in *Salm. typhimurium*. The work is not yet completed, but the data so far indicate that points not more than two nucleotides apart on the chromosome can be separated by recombination, and that one of the two functional loci will turn out to contain about 60 mutational sites.

Unfortunately the nature of the functional loss brought about by mutation to the r_{II} state has not yet been defined in biochemical terms. One of the most intriguing findings brought to

We can thus draw up a chromosome map in terms of time of transfer which has turned out, in fact, to be much more accurate than that arrived at by the more usual type of genetic analysis.

Relative distances between genes on donor chromosome as measured by:

	A	B	C	D	E
1. Mechanical fracture during transfer (blendor)	0·34	0·36	0·44	0·72	1·0
2. Decay of incorporated ^{32}P before transfer	0·35	0·39	0·43	0·67	1·0

FIG. 12. Relative distances between genes on the chromosome of *E. coli* measured (1) as a function of time by mechanical interruption of mating, and (2) as a function of ^{32}P decay in the donor chromosome before mating In each case, the distance of gene E from the extremity of the chromosome is taken as 1·0 (see Fuerst, Jacob and Wollman, 1956).

Now, if the DNA of *Escherichia coli* is made highly radioactive by incorporation of ^{32}P, it is found that the viability of the cells decreases exponentially as a function of ^{32}P decay. There is evidence to suggest that death is not due to radiation *per se*, but to fracture of the chromosome when ^{32}P atoms at the same point on opposite strands of the DNA double helix decay to ^{32}S (Stent and Fuerst, 1955). Fuerst, Jacob and Wollman (1956) examined the effect of ^{32}P decay in the DNA of donor cells on the ability of these cells to form recombinants. They found, as in the blendor experiment, that the further a gene is situated from the extremity of the chromosome, O, the more rapidly it disappears from recombinants as a function of ^{32}P decay.

When the relative rates of decrease for the various genes were compared with the relative times of entry of the same genes in a blendor experiment, they accorded extremely well as will be seen from Figure 12. What is being measured, then, in these experiments is the distance between genes in terms of the number of phosphorus atoms in the backbone of the DNA double helix. If the ^{32}P atoms are uniformly distributed along the length of the chromosome, then the further a gene is situated from O, the greater is the probability that a break due to ^{32}P decay will prevent its transfer to the zygote, just as when a blendor is used.

found that if the mating pairs were separated before eight minutes after mixing, no recombinants appeared; at eight minutes recombinants began to appear but these had inherited the donor gene A only (see Fig. 11). Shortly afterwards genes

E. COLI ZYGOTE

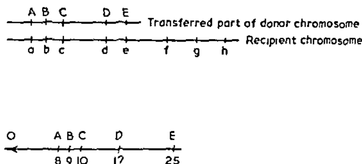


FIG. 11. The upper diagram represents the genetic constitution of a fertilized recipient cell (zygote) of *E. coli* Selection is usually made for recombinants inheriting gene A from the donor parent· the frequency of

lower diagram

B and C began to appear, in addition, among recombinants, to be followed at 17 minutes by D and at 25 minutes by E. The sequence with which these various genes appeared among recombinants was the same as the order of arrangement of the genes on the chromosome as determined by ordinary genetic analysis. The theory put forward to explain these facts is that the chromosome of the donor strain always penetrates the recipient cell by the same extremity, O, to be followed in turn by the other genes in the order of their arrangement on the chromosome. The effect of agitation is artificially to break the chromosome during transit; only those genes which have already entered the recipient cell at the time of treatment can participate in recombination and appear among recombinants.

We can thus draw up a chromosome map in terms of time of transfer which has turned out, in fact, to be much more accurate than that arrived at by the more usual type of genetic analysis.

Relative distances between genes on donor chromosome as measured by:

	A	B	C	D	E
1. Mechanical fracture during transfer (blendor)	0·34	0·36	0·44	0·72	1·0
2. Decay of incorporated ^{32}P before transfer	0·35	0·39	0·43	0·67	1·0

FIG. 12. Relative distances between genes on the chromosome of *E. coli* measured (1) as a function of time by mechanical interruption of mating, and (2) as a function of ^{32}P decay in the donor chromosome before mating. *In each case, the distance of gene E from the extremity of the chromosome is* taken as 1·0 (see Fuerst, Jacob and Wollman, 1956).

Now, if the DNA of *Escherichia coli* is made highly radioactive by incorporation of ^{32}P, it is found that the viability of the cells decreases exponentially as a function of ^{32}P decay. There is evidence to suggest that death is not due to radiation *per se*, but to fracture of the chromosome when ^{32}P atoms at the same point on opposite strands of the DNA double helix decay to ^{32}S (Stent and Fuerst, 1955). Fuerst, Jacob and Wollman (1956) examined the effect of ^{32}P decay in the DNA of donor cells on the ability of these cells to form recombinants. They found, as in the blendor experiment, that the further a gene is situated from the extremity of the chromosome, O, the more rapidly it disappears from recombinants as a function of ^{32}P decay.

When the relative rates of decrease for the various genes were compared with the relative times of entry of the same genes in a blendor experiment, they accorded extremely well as will be seen from Figure 12. What is being measured, then, in these experiments is the distance between genes in terms of the number of phosphorus atoms in the backbone of the DNA double helix. If the ^{32}P atoms are uniformly distributed along the length of the chromosome, then the further a gene is situated from O, the greater is the probability that a break due to ^{32}P decay will prevent its transfer to the zygote, just as when a blendor is used.

If it is assumed that the number of ^{32}P disintegrations required to prevent transfer is the same as the number required to kill a bacterium or a phage particle (which is known), then the number of phosphorus atoms and, therefore, the number of nucleotide units which separate the various genes can be estimated. It turns out that the length of chromosome transferred in one minute of time contains about 100,000 double nucleotide units. It is easy to calculate that if the whole chromosome were transferred at a uniform rate, this would take about 120 minutes, *giving about 12 million* double *nucleotide units* per *nucleus*. This figure is, in fact, about the same as that obtained by independent chemical estimations of the amount of DNA per nucleus.

In this lecture I have tried to show that the word 'gene', once so useful, is no longer very meaningful. As a result of recent work, which I have briefly reviewed, the tendency is to divide the chromosome into units of function, recombination and mutation, and to attempt to analyse these in molecular terms. In fact, Humpty Dumpty might have been referring to the gene when he said, in a rather scornful tone, 'When I use a word it means just what I choose it to mean—neither more nor less.'

REFERENCES

AVERY, O. T., MACLEOD, C. M. and MCCARTY, M. (1944). *J. exp. Med.* 79, 137.
BENZER, S. (1955) *Proc. Nat Acad Sci., Wash.* 41, 344.
BENZER, S. (1957). *The Chemical Basis of Heredity*, ed McElroy, W. D. and Glass, B., Baltimore, p. 70
BENZER, S and FREESE, E (1958) *Proc. Nat. Acad. Sci., Wash* 44, 112.
BRENNER, S , BENZER, S. and BARNETT, L. (1958) *Nature*, 182, 983.
CRICK, F. H. C., GRIFFITH, J S and ORGEL, L. E. (1957). *Proc Nat. Acad. Sci., Wash.* 43, 416
DELBRUCK, M. (1955) *Proc. Nat. Acad. Sci , Wash* 40, 783
DELBRUCK, M. and STENT, G. S. (1957) *The Chemical Basis of Heredity*, ed McElroy, W. D. and Glass, B , Baltimore, p. 699.
EPHRUSSI-TAYLOR, H. (1951). *Exp Cell. Res*, 2, 589.

FEUGHELMAN, M., LANGRIDGE, R., SEEDS, W. E., STOKES, A. A., WILSON, H. R., HOOPER, C W., WILKINS, M. H. E., BARKLAY, R. K. and HAMILTON, L D. (1955) *Nature*, **175**, 834.

FRAENKEL-CONRAT, H., SINGER, B. A and WILLIAMS, R. C. (1957). *The Chemical Basis of Heredity*, ed. McElroy, W. D. and Glass, B., Baltimore, p. 501.

FUERST, C. R., JACOB, F. and WOLLMAN, E. L. (1956). *C.R. Acad. Sci., Paris*, **243**, 2162.

GRIFFITH, F. (1928). *J. Hyg , Camb.* **27**, 113.

HARTMAN, P. E. (1956). Carnegie Inst. Wash Publ. no 612, 36.

HARTMAN, P E. (1957). *The Chemical Basis of Heredity*, ed McElroy, W. D. and Glass, B., Baltimore, p. 408

HERSHEY, A. D. and CHASE, M. (1952). *J. Gen. Physiol* **36**, 39

MESELSON, M. and STAHL, F. I. N. (1958). *Proc. Nat. Acad. Sci , Wash* **44**, 671.

STENT, G. S. (1955). *J Gen. Physiol.* **38**, 853

STENT, G. S and FUERST, C. R. (1955). *J. Gen. Physiol.* **38**, 441.

WATSON, J. D. and CRICK, F. H. C. (1953). *Cold Spr. Harb. Symp. Quant. Biol.* **18**, 123.

WOLLMAN, E L. and JACOB, F. (1955). *C.R. Acad. Sci., Paris*, **240**, 2449.

WOLLMAN, E. L. and JACOB, F. (1958). *Ann. Inst. Pasteur*, **95**, 641.

WOLLMAN, E. L , JACOB, F. and HAYES, W. (1958). *Cold Spr. Harb. Quant. Biol.* **21**, 141.

ZINDER, N D and LEDERBERG, J. (1952) *J Bacteriol.* **64**, 679.

VIII

Vaccination against Poliomyelitis

D. G. EVANS

IN 1955, when it was decided to embark on a programme of vaccination against poliomyelitis in Britain, the Medical Research Council's Biological Standards Control Laboratory at Hampstead was faced with the task of controlling all batches of poliomyelitis vaccine which were to be used in the immunization campaign. The task was divided into two main parts—the testing of vaccine for safety and freedom from living poliomyelitis virus and the testing for ability to produce adequate immunity. It is the second part of the task—the potency of the vaccine—which I wish to discuss.

PREPARATION OF POLIOMYELITIS VACCINE

A few of my readers may not be entirely familiar with some of the basic facts about poliomyelitis vaccine and it may be advisable first to give a brief outline of the method devised by Dr. Salk of preparing the prophylactic. It consists of the three types of poliovirus—types 1, 2 and 3. A strain of each type is grown separately in cultures of monkey kidney tissue in nutrient medium at 37° C.; the virus enters the tissue cells where it multiplies, breaks down the cells and pours into the nutrient fluid. After three or four days the fluid contains a high concentration of living poliovirus and is separated from the disintegrated monkey kidney by filtration. The filtered fluid is then subjected to the action of formaldehyde, usually at a concentration of 1/10,000 (1/4,000 formalin), at a temperature of 37° C. The formaldehyde is allowed to act on the virus for about 12 to 14 days, during which time all the virus particles

are killed. Halfway through the inactivation period it is usual to filter the suspension to remove aggregates of virus which may be resistant to formaldehyde. When killed suspensions of all three types have been obtained and shown to be safe and free from living virus, they are pooled together in suitable proportions to make a trivalent pool. The formaldehyde is then neutralized, preservative added and the vaccine distributed into ampoules and vials. Further tests for safety and potency are made on the final product—these tests usually take about ten weeks to complete—and when they have been shown to be satisfactory the vaccine is ready for use.

The virus suspension at the beginning of the inactivation period contains more than one million particles of living virus per ml. and slowly, under the action of formaldehyde at 37° C., the number is reduced and at 6 days tests made with large volumes are usually unable to detect live virus. A longer period, however, is employed for inactivation to increase the safety of the vaccine. During the period of inactivation there is also a fall-off in the immunizing activity of the virus suspension but this is by no means as rapid as the fall-off in viability.

Variables in Vaccine Production

One might justifiably assume that the method I have described of preparing poliomyelitis vaccine would produce batches of vaccine with a consistent level of antigenic activity. This, however, is by no means the case, for although the principle of preparing the vaccine is simple, the actual operations are most complicated and involve a considerable number of variable factors.

In the first place, the strains may vary. In this country we are using the avirulent Brunenders strain for type 1, strain MEF-1 for type 2 and Saukett strain for type 3; in the United States of America and Canada the same strains are used for types 2 and 3 but for type 1 the virulent Mahoney strain is employed while other countries have chosen their own particular strains. The tissue culture preparations vary—some manufacturers use minced monkey kidney suspended in the culture fluid while others use trypsinized tissues which give cultures of cells growing in layers in the culture vessels. The culture medium is also a

variable factor and although the most popular one is Parker 199, there are, nevertheless, numerous modifications of this. The method of inactivation varies from manufacturer to manufacturer; formaldehyde is widely used but other agents, such as ultraviolet light and β-propiolactone, are sometimes employed in conjunction with formaldehyde.

Time and temperature of inactivation are also variable factors. Most manufacturers inactivate at 37° C. for 12 to 14 days but occasionally a lower temperature (25° C.) for a longer period is adopted since it is considered that this method has a less destructive effect on the antigen. A major variable during production is the filtration process which takes place halfway through the inactivation period. This filtration process has been a subject of much controversy and there seems, as yet, no accepted standard method. Two main types of filter are employed—the Seitz pad and the sintered glass filter. It is well known that each type may vary considerably and furthermore there are various ways in which the filters are treated before the virus suspension is passed through them. It is extremely difficult to forecast the amount of virus adsorption which may occur and what effect filtration may have on antigenic activity.

There is no question that these and many other factors have an effect on the antigenic activity of the final product and it is not surprising, therefore, to find that different vaccines made by different methods, and in some cases even by the same method, vary considerably in their activity.

Test for Antigenic Potency

Before I discuss the results which show how potency may vary from vaccine to vaccine, I should perhaps describe the way in which the antigenic activity of a vaccine is tested in the laboratory. The method employed is one which was laid down by the United States National Institutes of Health in the early days of poliomyelitis immunization. It involves the immunization of a group of 12 monkeys, each animal being given three doses, each of 1 ml., at intervals of one week. Blood samples are taken 7 days after the third dose and tested to determine the concentration of neutralizing antibodies to all three types of poliovirus.

There are numerous ways in which the neutralization test may be made, but I would like to outline the one used routinely by the Biological Standards Control Laboratory (1957). It is similar in principle to many other methods of measuring antibody concentration in serum. A series of dilutions of the serum is first made and to each dilution a constant amount of living poliovirus is added. Serum and virus are thoroughly mixed and incubated at 37° C. for three hours. This period and temperature of incubation were shown by Perkins, Placido Sousa and Tobin (1958) to be essential in order to allow antibodies to react completely with virus; with shorter periods unsatisfactory results are obtained on account of incomplete neutralization. After the incubation period, the mixtures are inoculated into tubes containing actively growing monkey kidney cell cultures and allowed to stay in contact with the cultures for 6 days. If the virus has been neutralized by the serum the cultures show, on microscopic examination, a homogeneous pattern of healthy cells. If, however, the concentration of antibodies is too small the virus remains un-neutralized and attacks the cell culture causing extensive degeneration. This method gives a measure (or titre) of the antibody concentration in the serum, the titre being expressed as the highest dilution of serum which is able to neutralize the virus suspension. It is hoped that soon we shall be able to abandon this unsatisfactory method of expressing antibody concentrations by employing standard antisera and expressing potency in terms of units. Standard antisera have already been prepared and will be put into general use in this country when a unitage has finally been assigned to them.

Variation in Potency of Vaccines

Employing this method of testing antigenic activity by immunizing monkeys with vaccine and titrating their sera, we have found considerable differences in potency from vaccine to vaccine. Figure 1 shows the results of potency tests made on 38 different vaccines. The results are given for each of the three virus types, the potency being expressed as the logarithm of the reciprocal of the mean antibody titre obtained with a batch of 12 monkeys. For each vaccine there are three points, one for

each type, each point representing the level of potency for the particular type. It is evident that, with all three types, there are considerable differences between vaccines, some being almost a hundred times as potent as others.

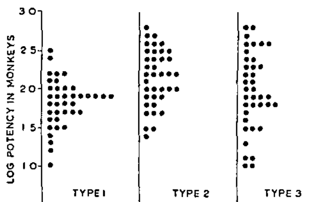


FIG. 1. Potency of 38 batches of poliomyelitis vaccine, determined in monkeys.

RELATION BETWEEN LABORATORY AND FIELD TESTS

We must now consider the question of which vaccines are able to produce an adequate response in humans. In other words, can a line of demarcation be drawn on the basis of laboratory tests between those vaccines which are good immunizing agents in humans and those which are poor? This is a problem which has to be faced in controlling the potency of all prophylactics. The problem arose when the diphtheria immunization campaign was launched in this country in 1940 and it was shown then, by comparing the potency of different batches of alum precipitated diphtheria toxoid in guinea-pigs with their ability to immunize children, that the guinea-pig test was able to distinguish between those toxoids which gave poor and those which gave good protection to children. The same problem arose with pertussis vaccine. In this case the Whooping Cough Immunization Committee of the Medical Research Council, by carrying out extensive laboratory and field tests, showed that it was possible, by using a mouse-protection test, to distinguish between those vaccines which were able to give good protection

in children and those which were not (1956). Can the same be done in the case of poliomyelitis vaccine? Can we, by testing the antigenic potency of a vaccine in monkeys, decide whether it will give an adequate antibody response in humans? This question has been investigated by the Biological Standards Control Laboratory, in collaboration with field workers, by comparing the antigenic activity of a number of vaccines in monkeys with their activity in humans.

The tests in monkeys were quite straightforward—they were done by the method I have already outlined. The tests in humans, however, were more difficult since it was necessary to obtain a sufficiently large number of non-immune human subjects. We were extremely fortunate in having the co-operation of three large schools—Epsom College, Mill Hill School and Aldenham School. All the boys who took part in the study first had their sera titrated for antibodies against poliomyelitis. Those boys who had no antibodies to any of the three types were specially selected and given two doses of vaccine at a month's interval. Two weeks after the second dose, second samples of sera were taken and these also were titrated for antibodies. From the results obtained we were able to determine, for each vaccine on test, the percentage of boys who responded to each type of poliomyelitis virus, and this was compared with the antigenic activity of the vaccine as determined by the monkey test. This work is by no means complete but some results are available and are given in Figure 2, which shows the comparison of field and laboratory results for eleven vaccines.

It is seen that 90 to 100 per cent of boys responded to types 2 and 3 with all vaccines tested. This was rather unfortunate from the point of view of correlating laboratory and field results, but, nevertheless, it can be concluded that if a vaccine gives in monkeys a log-potency of 2 or greater, then it will give a substantial response in humans. When we turn to type 1, however, the picture is somewhat different. Here we were more fortunate—from the experimental standpoint—in testing vaccines of both low and high activity, and because of this we were able to show a certain degree of correlation between the activity in humans and that in monkeys. It is quite clear that those

vaccines which give in monkeys a log-potency of 2 or more were able to induce an antibody response in a high percentage of humans, but as the potencies of the vaccines in monkeys fell below this level, a rapid fall occurred in their immunizing activity in the field.

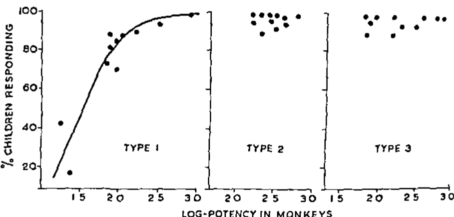


FIG. 2. Relation between the results of field and laboratory tests with 11 poliomyelitis vaccines.

These results enable us to use a method of controlling potency and of ensuring that only potent vaccines are issued for use in the field. If we now return to Figure 1, it can be seen how this method may be applied. If a line is drawn across this diagram at the level of log 2, it can be said that those vaccines with a potency above this level are satisfactory but that those below are doubtful in the case of types 2 and 3 and definitely unsatisfactory in the case of type 1. This criterion could hardly have been applied to the vaccines used during the last three years—in the first place it has taken until now for the method to be *investigated and, secondly, vaccine has been in short supply* and it was better to use vaccine which would immunize at least some children than no vaccine at all. But now we are approaching a much more favourable position when vaccine will no longer be in short supply and in such a situation we need choose only the best vaccines by applying this system of control. One might suppose that the system will be rather severe and rule out

the use of many batches of vaccine. This, however, will not be the case for already vaccines well above this line of demarcation for all three types are being produced by British manufacturers.

Although this method of assaying vaccines is extremely valuable, it has one major defect. As it stands it is not biometrically sound since no standard vaccine is used for comparison. In the ideal biological test a stable standard preparation must be employed as a 'yardstick' in order to eliminate the test variations which occur from day to day. At present there is no standard preparation of dry stable poliomyelitis vaccine although efforts are being made in many countries to prepare one. Until such a standard vaccine is obtained we shall have to depend on the test as it is done at present, and although this is not ideal it will, nevertheless, draw a broad distinction between poor and good vaccines. Actually the control of diphtheria and tetanus toxoids has been carried out for many years without the use of standard preparations for comparison, but with both these toxoids the level of potency demanded by the control authorities, a level which manufacturing establishments have well exceeded, has resulted in the production of prophylactics of high immunizing quality. There is, thus, every reason to believe that a similar situation will result in the case of poliomyelitis vaccine.

The Duration of Immunity after Vaccination

Although this is a decided step forward in the laboratory control of poliomyelitis vaccines and in ensuring that vaccines used in the field are above a certain level of antigenic activity, there is another aspect of the problem which must be considered. So far I have only referred to the immune response immediately following two doses of vaccine, whereas in vaccinating against poliomyelitis we are not merely concerned with the immediate response to two doses but rather with the lasting effect of the immunity. It is well known that a more substantial immunity, and one of longer duration, can be obtained in prophylactic procedures—such as diphtheria and tetanus—by giving a third or 'booster' dose a year after the primary two doses and this principle also applies to vaccination against poliomyelitis. It must be remembered, however, that for the third dose to be

effective and act as a booster, a sufficient basal immunity must first be established by primary immunization. Thus, a further question is raised—do those vaccines which give, after two doses, an antibody response in a high proportion of subjects *also* induce a sufficient basal immunity to enable the third dose to act as a booster?

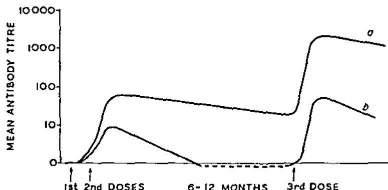


FIG. 3. Antibody response of schoolboys to poliomyelitis vaccine.

This question has been investigated by Dr. J. O'H. Tobin and his colleagues at Hampstead in collaboration with Dr. J. C. Kendall at Epsom College (1959) with groups of non-immune schoolboys. The boys were first given two doses of vaccine at a month's interval and their antibody levels determined two weeks after the second dose. At 6 to 12 months their antibody levels were again tested and they were then given a third dose of vaccine. Further samples of sera were taken two weeks after the third dose and these also were tested for antibodies. The results of this study are shown in Figure 3. With types 2 and 3 the antibody response followed the course indicated by line *a*; the antibody level resulting from the two primary doses was not completely lost after 6 to 12 months and the third dose of vaccine produced a striking increase in titre with both types. This is a very satisfactory picture and bears a close resemblance to that which occurs in immunization against tetanus and diphtheria, and it is reasonable to suppose that the immunity produced from this three-dose schedule will be of considerable

duration. With type 1, however, a different result was obtained. Although two doses gave a response in a high proportion of subjects, the group was divided into about 70 per cent who responded well, following line *a* in Figure 3, and the rest who responded slightly less, following line *b*. The first group behaved as with types 2 and 3—they retained their basal immunity during the following 6 to 12 months and gave the typical and satisfactory booster response to the third dose. The other group —those who responded less well—lost their basal immunity during the 6- to 12-month period and when given the third dose the response, instead of being of the booster type, was of the primary type, the level of antibody being of the same order as that obtained after two doses and obviously one which will not be long-lasting in its effect.

These results strengthen the view, which I have already expressed, that we must aim at preparing vaccines of high antigenic quality. It must be borne in mind that vaccines must be such that they will give not merely a primary response in humans, *but a primary response which is substantial enough to ensure that all subjects will give a booster response to the third injection.* The majority of batches of vaccine produced at present come up to this standard with regard to types 2 and 3 but many of them fall below it with regard to type 1. It is absolutely essential that efforts should be made to remedy this defect, since immunity to

the proportion of the three components of the vaccine; work along this line is in progress at the Biological Standards Control Laboratory and at the manufacturing establishments in this country and should ultimately result in the production of poliomyelitis vaccine with sufficient activity to give an immunity of long duration to all three types.

THE RELATIONSHIP OF VACCINATION AGAINST POLIOMYELITIS TO OTHER IMMUNIZATION PROCEDURES

There is one further question I would like to discuss. How does immunization against poliomyelitis fit in with immunization

against other diseases such as diphtheria, tetanus and whooping cough? During the last ten to fifteen years there has been an increasing tendency to used mixed prophylactics, and vaccine containing diphtheria, pertussis and tetanus antigens—triple

Fig. 4. Antibody response to poliomyelitis vaccine of infants of different ages.

vaccine—has been used on an increasing scale for immunization in early infancy. Triple vaccine has the great advantage of reducing the number of injections which the child receives. The

...rovoking effect
result of these
it possible to
immunize children against poliomyelitis early in infancy and
...ccine

and Gaisford (1958) by studying the response of young infants to two doses of poliomyelitis vaccine. They found that infants at one week after birth gave a very poor response, that a slightly better one was given at 6 weeks and still slightly better at 10 weeks, but by no means as complete as with 12-month-old children. This was the case with each virus type (Fig. 4). The inability to respond was shown to be due to the interference of

placentally transmitted antibodies, an effect which wears off with age as the antibodies are eliminated. As a result of this study, it was concluded that it was better to wait before giving poliomyelitis vaccine until the child was at least 6 to 9 months of age when little or no interference from maternal antibodies would occur. Thus, if triple vaccine is to be given under the

given early in infancy and that 7 months after birth is too long to wait. There is one way of approaching this difficulty—namely, to immunize the mother against poliomyelitis during pregnancy. It will then be possible to give the child triple vaccine under the cover of placentally transmitted antibodies during the first few months of life without the risk of provoking poliomyelitis. This would seem a reasonable and desirable procedure, for not only would it protect the child against the provoking effect of triple vaccine, but it would also protect the pregnant mother against poliomyelitis—for it must be remembered that there is a higher incidence of the disease during pregnancy. The general scheme would be as shown in Figure 5, which gives the poliomyelitis antibody levels of the mother during pregnancy and of the child during the first few years of life. The level of antibody produced by immunization during pregnancy would, of course, depend on whether the mother had a basal immunity to poliomyelitis as a result of either earlier immunization or an earlier history of poliomyelitis. If so, then the child would be born with a very high level of poliomyelitis antibodies. In any case, the maternal antibodies would be present until the child was 4 months of age and during this period, at say 2, 3 and 4 months or earlier if possible, three injections of triple vaccine could be given without risk. Later, at 8 to 9 months, when the maternal antibodies had completely disappeared, the child could receive its two primary inoculations of poliomyelitis vaccine which would be followed a year later, at the age of 20 months, by a booster dose of poliomyelitis vaccine and at 2 to 3 years of age by a booster dose of triple vaccine.

It may be possible in time to combine the booster dose of

against other diseases such as diphtheria, tetanus and whooping cough? During the last ten to fifteen years there has been an increasing tendency to used mixed prophylactics, and vaccine containing diphtheria, pertussis and tetanus antigens—triple

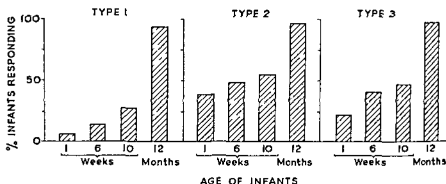


FIG. 4 Antibody response to poliomyelitis vaccine of infants of different ages.

vaccine—has been used on an increasing scale for immunization in early infancy. Triple vaccine has the great advantage of reducing the number of injections which the child receives. The use of mixed prophylactics, especially those containing alum, had a serious setback when it was shown by a Medical Research Council Committee (1956) that there was a definite risk in

oking effect

ılt of these

possible to

immunize children against poliomyelitis early in infancy and

Gaisford (1958) by studying the response of young infants to two doses of poliomyelitis vaccine. They found that infants at one week after birth gave a very poor response, that a slightly better one was given at 6 weeks and still slightly better at 10 weeks, but by no means as complete as with 12-month-old children. This was the case with each virus type (Fig. 4). The inability to respond was shown to be due to the interference of

for England and Wales for pertussis, diphtheria and tetanus in relation to the age of the child (Fig. 6). Here it is seen that the great majority of deaths from pertussis infection occur within the first year of life, that between 1 and 5 years of age there is a

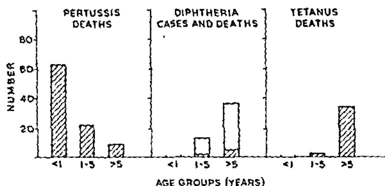


Fig. 6. Deaths from pertussis and tetanus, and cases and deaths from diphtheria in different age groups in England and Wales during 1956.

definite reduction and that over 5 years a still further reduction [illegible]

tetanus, however, the picture is different. With both these diseases no cases or deaths occurred in the first year of life, a small number in the period from 1 to 5 years and the maximum number in the group over 5 years of age. It would, therefore, seem reasonable to separate the pertussis inoculations from those of diphtheria and tetanus and follow a scheme as outlined in Figure 7.

In this scheme three injections of pertussis vaccine—which when unmixed with other prophylactics has relatively no provoking effect—could be given at 2, 3 and 4 months. This would be followed at 8 and 9 months by two primary doses of poliomyelitis vaccine and a year later by the booster dose. When the child is entering into the danger zone of diphtheria and tetanus, say at 2 years, two primary doses would be given of mixed diphtheria and tetanus toxoids, with alum if desired, since the

poliomyelitis vaccine with that of the triple vaccine and thus further decrease the number of injections the child receives. The use of a quadruple vaccine—containing poliomyelitis, diphtheria, tetanus and whooping cough antigens—should,

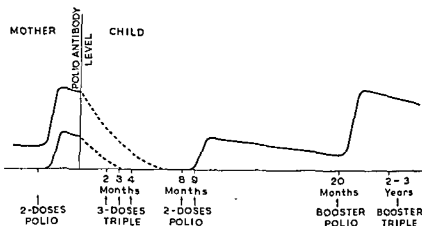


FIG 5. Poliomyelitis antibody level in children in relation to immunization schedule.

however, *not* be embarked upon until we are quite certain that we have a high-quality poliomyelitis vaccine. For if poliomyelitis vaccine of borderline potency is combined with strong antigens such as diphtheria and tetanus toxoids there is every likelihood that the stronger antigens will suppress the weaker ones and especially will this be true at the booster stage when the child possesses a basal immunity.

There are two possible disadvantages with this scheme of immunization. In the first place, it depends entirely upon whether the mother during pregnancy will accept poliomyelitis vaccine, and there is evidence from the response so far obtained in this country that she is extremely reluctant. In the second place, it is known that diphtheria and tetanus toxoids when given early in life do not produce as good an antigenic stimulus as when given later. It is, therefore, worthwhile considering other schemes of immunization and in doing so I have considered them in relation to the incidence of the various diseases against which we are immunizing. Let us look at the 1956 figures

REFERENCES

BIOLOGICAL STANDARDS CONTROL LABORATORY (1957). 'Safety and Antigenic Potency Testing of Poliomyelitis Vaccine.' *Brit med. J.* ii, 124.

MEDICAL RESEARCH COUNCIL (1956). 'Vaccination against Whooping-Cough.' *Brit. med J.* ii, 454.

MEDICAL RESEARCH COUNCIL (1956). 'Poliomyelitis and Prophylactic Inoculations against Diphtheria, Whooping-cough and Smallpox.' *Lancet*, ii, 1223.

PERKINS, F. T., PLACIDO SOUSA, C. A. and TOBIN, J. O'H (1958). 'The Titration of Poliomyelitis Neutralizing Antibodies.' *Brit J. exp. Path.* 39, 171.

PERKINS, F. T., YETTS, RISHA and GAISFORD, W. (1958). 'Serological Response of Infants to Poliomyelitis Vaccine' *Brit. med. J* ii, 68

TOBIN, J. O'H. (1959). Report on 'Antibody Response of Adolescents and Adults to a Booster Dose of Poliomyelitis Vaccine'. *Brit. med. J.* i, 609.

child will be immune to poliomyelitis, and followed up with a booster dose of mixed diphtheria and tetanus toxoids at 4 to 5 years—the child would then have a substantial immunity as it entered school life.

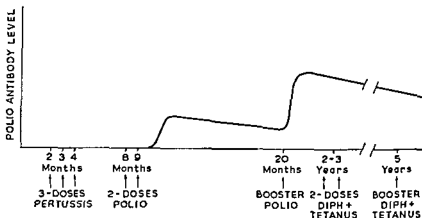


Fig. 7. Poliomyelitis antibody level in children in relation to *immunization schedule.*

Both these immunization schemes have certain advantages but the first appeals to me more than the second. In the first

out risk, with triple vaccine against whooping cough, diphtheria and tetanus. By the age of 12 months it has developed, as a result of five injections, an active basal immunity to four diseases and two more injections during the following 12 months give the child an immunity of long duration to these diseases. Altogether the scheme involves only seven inoculations. It is a scheme I would like to see investigated more fully and, if it should prove safe—and I see no reason why it should not—one I would like to see adopted throughout the whole country. For such a scheme would ensure—with a minimum number of inoculations—a high degree of protection in childhood against diphtheria, tetanus, whooping cough and poliomyelitis.

but almost at once another virus resembling it closely was reported from the Army Medical School, Washington, by Hilleman and Werner (1954). These authors cultivated throat washings from cases of acute respiratory disease occurring in the U.S. Army using the Hela strain of human cervical carcinoma as medium. Cytopathic effects were demonstrated in the cells and neutralization and complement-fixation techniques were successfully applied to human sera using antigens prepared from the virus culture. This second strain of adenovirus called by Hilleman and Werner the RI-67 agent was shown to be a cause of human respiratory disease and a close serological link with the adenoid-degenerating agent was soon reported. For some time developments occurred largely in the U.S.A. where various authors recovered strains of similar viruses and gave each of them separate names including A.P.C. or adenoidal-pharyngeal-conjunctival virus. By mutual consent, however, a new terminology was introduced in 1956 and the term adenovirus then came into use and has replaced all earlier names (Enders *et al.*, 1956).

Characters of the Adenoviruses

The adenoviruses are now recognized as a widely prevalent group of agents existing in the form of multiple antigenic types, of which 18 have now been recovered from man and 5 additional types from monkeys or chimpanzees (Rowe *et al.*, 1958). All types share a common complement-fixing antigen which appears in tissue cultures infected by the virus. It is known as the 'soluble' complement-fixing antigen, and it is used extensively in the investigation of human infection by the demonstration of rises in antibody titre accompanying clinical events. Whenever recognition of the exact virus type causing infection is needed, a neutralization test based upon tissue culture methods is necessary. Neutralization tests are also required for the identification of strains of adenoviruses recovered from human specimens. The situation is somewhat similar to that existing in the influenza viruses and this analogy is increased by the discovery of haemagglutinins of a type-specific character in adenovirus cultures (Rosen, 1958).

The adenoviruses do not produce pathological effects when

IX

The Adenoviruses and Respiratory Disease in Man

C. H. STUART-HARRIS

The Recovery of the Adenoviruses

THE discovery of the adenoviruses in 1953 was one of the many important by-products of the new tissue culture technique for the cultivation of viruses introduced by Enders and his colleagues at the Boston Children's Hospital. Although Feller, Enders and Weller first grew vaccinia virus in chicken embryo cultures in test-tubes in 1940 the technique achieved its first outstanding success in 1949 when Weller, Robbins and Enders cultivated poliomyelitis virus in human tissue cultures. Their roller tube method permitted easy recognition of the growth of viruses such as poliovirus, which produce striking destructive effects upon cells, and it also enabled cultures of cells to be kept alive for many days during which the medium was changed and all traces of antibodies in the original tissue were removed. This technique of prolonged cultivation was applied by workers at the National Institutes of Health, Bethesda, to human tissue removed surgically. Rowe and his co-workers found in 1953 that, after prolonged cultivation, degeneration occurred in the epithelial cells growing out from fragments of adenoids removed from children, and that these cytopathic effects were transmissible to other healthy tissue cultures of human origin. Huebner describes this technique as one of 'unmasking' viruses or of permitting viruses to 'crawl out' of cells in which they are harboured in a phase of inapparent or latent infection. Thus the first strain of adenovirus was obtained,

and Ginsberg (1957) demonstrated that they contained deoxyribonucleic acid. The packing of the bodies into a crystalline pattern has attracted much attention and may be explained by the polyhedral shape of the individual particles recently demonstrated by Valentine and Hopper (1957) at the National Institute for Medical Research. Morgan and others (1957) have also reported the presence of a second type of intranuclear body in cells infected with Type 5 adenovirus and this seems to be a protein. An interesting analogy with non-infective crystalline protein found in plant cells infected with tobacco mosaic virus is thus raised.

These then are the main characters so far established for this group of viruses, some strains of which are certainly widely distributed in man. There are probably many other interesting properties still to be discovered on which it would be nice to speculate but the task set in this lecture is to consider the relationship of the virus to respiratory disease in man.

The Syndromes of Adenovirus Infection

(a) Febrile catarrh

It is now necessary to look backwards into the history of the development of knowledge of the acute virus respiratory diseases in man. In 1935 soon after the recovery of the first strains of influenza virus in the laboratory by the inoculation of ferrets, workers at the National Institute for Medical Research studied an outbreak of acute respiratory disease in soldiers at Woolwich. This, though resembling influenza clinically, was not associated with the ferret-pathogenic virus (Andrewes, Laidlaw and Smith, 1935). Correlated clinical and laboratory studies were therefore made in 1936 of four outbreaks, three of which occurred in military establishments and one in a public school. Influenza virus was not recovered from any of the patients thus studied and no serological evidence was obtained suggesting that influenza virus was concerned. The widespread epidemic of influenza in 1936–7 which readily gave evidence of influenza virus infection was also studied in military communities and the clinical and epidemiological findings in this outbreak were

inoculated into animals such as mice, hamsters, rabbits or fertile hens' eggs. They grow well in tissue cultures of human epithelial cells either from normal or malignant tissue and some strains can be induced to produce cytopathic changes in rabbit kidney tissue (Balducci, Zaiman and Tyrrell, 1956; Kelly and Pereira, 1957) although there is only limited virus multiplication. Degenerative changes are also produced by adenoviruses in monkey kidney cultures but these cells are not such a good medium as are human cultures. Certain subtypes of adenoviruses have been found in monkey and chimpanzee specimens but the possibility exists that these hosts acquired the viruses as a result of human contact. In man the viruses have been recovered from respiratory tract material such as throat washings, adenoids and tonsils, and autopsy or resected lung tissue. They have also been found in faeces and in mesenteric lymph nodes. Their pathological effects in tissue cultures consist of a rounding-up of cells into clumps which fall from the walls of the tube into the supernatant fluid. The cells are not killed, however, but continue to undergo metabolism and considerable amounts of acids such as lactic, pyruvic, acetic and alpha-ketoglutaric acid accumulate in the medium (Fisher and Ginsberg, 1957). This effect is strikingly different from that produced by other viruses which usually inhibit the metabolism of cells in which they grow. Because of their peculiar host-cell relationship the adenoviruses are thus well adapted to a symbiotic existence and this may account for their role of latent infection already mentioned. Even in a non-susceptible animal such as the rabbit, Pereira and Kelly (1957a) demonstrated the persistence of a strain of adenovirus in the spleen for at least two months after inoculation. [illegible] of spleen

[illegible] hat infected cells develop characteristic intranuclear bodies. Kjellén and co-workers (1955) first obtained electron microscope photographs of these bodies and showed that they consisted of regular [illegible] in size. Councilman

[illegible]

was apparently distinct from that of primary atypical pneumonia which could also be transmitted to volunteers with filtrates from appropriate cases of atypical pneumonia (Commission on Acute Respiratory Disease, 1946, 1947b and c). No fresh work was reported on the A.R.D. syndrome, except further confirmation that it was unrelated to the influenza viruses, until Hilleman and Werner (1954) recovered their strain of adenovirus Type 4 from cases of A.R.D. during an outbreak at Fort Leonard Wood in 1953. Dascomb and Hilleman (1956) later described the clinical features of proven adenovirus infection in 45 recruits at Fort Dix and agreed that these were similar to the Commission's syndrome of A.R.D. and an earlier account of febrile catarrh brought up to date in 1953 (Stuart-Harris, 1953). Comparative frequency of various symptoms and signs is listed in Table 1 which shows the three sets of observations. It seems

TABLE 1. Percentage frequency of symptoms in febrile catarrh, A R D. and adenovirus infection

Symptoms	Febrile Catarrh*		American A R.D.	Adenovirus†
	28 cases	17 cases	113 cases	45 cases
	1936	1945	1942–5	1954
(a) Respiratory				
Cough	86	88	85	90
Sore throat	80	70	69	100
Nasal symptoms	70	70	60	70
Expectoration	50	23	54	40
Hoarse voice	40	40	75	35
Pain in chest	40	10	48	50
(b) Constitutional				
Malaise	82	23	53	90
Shivering	70	53	69	40 (chills)
Anorexia	70	23	55	35
Headache	64	76	65	40
Dizziness	64	23	—	30
Muscular pains	53	30	—	42
Sweating	50	—	—	—
Insomnia	40	17	—	—
Abnormal signs chest	33	23	27	67
Average duration of fever	*3 days*	*?*	*2–3 days*	*5·5 days*

* Modified from Stuart-Harris (1953) † Dascomb and Hilleman (1956).

contrasted with those of the earlier outbreaks. Although we had no evidence that the non-influenzal outbreaks of 1936 were similar to each other in causation we drew together common findings and attempted to contrast these with virus influenza (Stuart-Harris, Andrewes and Smith, 1938). Because the non-influenzal disease was febrile and because local signs of inflammation of the upper respiratory tract were often prominent and even included exudate upon the throat in some patients, the name 'febrile catarrh' was used as a group description. This syndrome of febrile catarrh, though sometimes appearing in a form closely resembling influenza, was at other times a pharyngo-laryngotracheitis in which sore throat, hoarse voice and a painful, paroxysmal, irritating cough dominated the symptomatology. In still other patients the illness was best described as a feverish cold.

The use of this term *febrile catarrh* was not always understood by later observers. Although we intended it to be a word to cover clinical illnesses often occurring in epidemics which were not due to influenza virus, some have used it to cover the general mass of all forms of acute respiratory disease, including influenza. However, during the second world war, Army outbreaks of respiratory disease were studied in the U.S.A. by a commission working under the Armed Forces Epidemiological Board. This Commission on Acute Respiratory Diseases recognized that there existed in military establishments and particularly in those dealing with recruits, a common endemic disease not due to either of the serologically distinct influenza viruses A or B or to recognizable bacterial causes. The patients exhibited the various clinical illnesses which we had described as febrile catarrh. The Commission termed the syndrome undifferentiated acute respiratory disease or A.R.D. (Commission on Acute Respiratory Diseases, 1947a). They recognized that some patients exhibited an acute pharyngitis with exudate on the tonsils and that others showed an abnormal radiological picture in the chest termed atypical pneumonia. Moreover, the Commission established that the A.R.D. syndrome could be transmitted to human volunteers by inoculation with bacteria-free filtrates of nasal and pharyngeal washings, and that the cause

with rhinitis, pharyngitis and conjunctivitis. An adenovirus Type 3 was recovered from both eye and throat swabs of some of the patients and rises in neutralizing and complement-fixing serum antibodies were demonstrated in 8 instances. This report was followed by an account of outbreaks of fever with pharyngitis and conjunctivitis occurring in children's camps in the vicinity of Washington in the summer of 1954 (Bell *et al.*, 1955). Type 3 adenovirus was incriminated and the syndrome was designated pharyngo-conjunctival fever. The chief difference from febrile catarrh was the conjunctivitis which was either unilateral or bilateral and of the type termed follicular; it occurred in nearly 60 per cent of cases. A moderate degree of cervical or occasionally pre-auricular lymph gland enlargement was also noted. The outbreaks were thought to have been spread by swimming-baths.

Since then outbreaks of a similar character and associated with adenoviruses occurred among children during 1955 in Helsinki (Forssell *et al.*, 1956; Oker-Blom *et al.*, 1957), in Sweden (Kjellén, Sterner and Svedmyr, 1957), and in Britain. The types of viruses recovered from the outbreaks varied. The Finnish workers and my own group in Sheffield (Tyrrell, Balducci and Zaiman, 1956) recovered the variant Type 7 virus already mentioned. The outbreaks in public schools described by Kendall and others (1957) and which occurred in Britain in 1955 yielded Type 3 in three instances and Types 7 and 14 in one outbreak each. The clinical features resembled those described in the U.S.A. but conjunctivitis was variable in occurrence. In a more recent outbreak of pharyngitis in a boys' school in February 1957 conjunctivitis was exceptional although Type 3 virus was incriminated (Munro-Ashman *et al.*, 1958).

In Sweden Type 3 virus was recovered from an epidemic of pharyngo-conjunctival fever (Kjellén, Zetterberg and Svedmyr, 1957) and also from scattered cases of pharyngitis admitted to hospital in Stockholm. Many of those who have reported outbreaks of pharyngitis associated with adenovirus infection have commented on the occasional occurrence of gastro-intestinal symptoms including abdominal pain, vomiting or diarrhoea. A

likely that Dascomb and Hilleman's patients were more severely ill than the earlier patients as shown by the duration of fever and percentage with signs of tracheobronchitis. Ginsberg and co-authors (1955) have since published serological work with stored sera from wartime cases of A.R.D. induced in volunteers which established that the causative agent of the filtrate disease was an adenovirus.

In Britain evidence was found that adenovirus infection was present in military populations in the North of England in 1955 by Tyrrell, Balducci and Zaiman (1956) and elsewhere in England by Andrews, McDonald, Thorburn and Wilson (1956) working in the R.A.F. It appears that the infection in Britain was chiefly due to a serological variant of adenovirus Type 7 whereas in the U.S.A. Army outbreaks and in others occurring in the Navy (Woolridge *et al.*, 1956; Rowe *et al.*, 1956), Types 4 or 7 were chiefly responsible. In France, Sohier and others (1957) also found the variant Type 7 to be a cause of an outbreak among soldiers at Lyon. In Holland, on the other hand, Van der Veen and Kok (1957) investigated an outbreak among recruits which was found to be due to Type 14 of which the prototype strain was originally recovered from a chimpanzee. Time alone will show whether future outbreaks of febrile catarrh are associated with other antigenic types as may, of course, be anticipated. A warning is perhaps necessary that many observers have found that only a proportion of cases of febrile catarrh in military establishments yields evidence of adenovirus infection. Thus in the R.A.F. McDonald and others (1958) found almost no cases of adenovirus infection in 1956–7 in recruits, though the same establishments had experienced sharp outbreaks in the previous year. The causative agent of much acute respiratory disease in the Services has yet to be established.

(b) Pharyngo-conjunctival fever

Towards the end of 1954 some twelve months after the recovery of the first strain of adenovirus, workers at the National Institute of Health (Parrott *et al.*, 1954) published an account of a ward epidemic of febrile pharyngitis occurring at the Clinical Centre in Bethesda, following upon the admission of a two-year-old girl

reported from Canada (Fowle, Cockeram and Ormsby, 1955), the U.S.A. (Jawetz *et al.*, 1956), Japan (Tanaka, 1958) and Britain (Bennett *et al.*, 1957; Sommerville, 1957). It appears that Type 8 is particularly responsible for such outbreaks though other types of adenovirus (1, 7 or 11) have been recovered from time to time. Shipyard workers have been particularly affected in Britain and elsewhere.

(e) Respiratory disease in the general population

In spite of the indisputable evidence of adenovirus infection derived from outbreaks in semi-isolated communities such as servicemen and public schools, it has proved difficult to associate the viruses with the common respiratory disease of the general population. Difficulties arise, of course, in any attempt to relate specific infectious agents to clinical events in the community as a whole because there is no regular system of surveillance of morbidity in the latter. There are two possible methods of approach. The first method is to conduct a survey of the distribution of antibodies in random samples of the population. This is easy to achieve with viruses existing in a single antigenic form but much more complicated when antigenic variation exists. Nevertheless serological studies have been made in the U.S.A. which reveal quite precisely the speed of acquisition of antibodies to the first 7 types of adenoviruses. Wenner and co-authors (1957) studied a group of 88 families and showed that antibodies against Types 2 and 5 adenoviruses developed early in infancy. Against Type 4 on the other hand antibodies were slow to appear and even adult sera frequently showed no neutralizing power against this virus. Jordan, Badger and Dingle (1958) who have studied a group of families in Cleveland for a long period of years found that Types 1 and 2 infections were common during the first five years of life but

antibodies to Types 3, 4 and 7.

When the reason for the prevalence of antibodies to Types 1 and 2 in early infancy is sought, however, it is clear that most of

good example of this is a recent study of 38 cases of Type 3 infection described by Barr and others (1958). It is therefore interesting to note that adenoviruses have been recovered from faeces as well as from the throat during attacks of pharyngitis (Kjellén, Zetterberg and Svedmyr, 1957). Swedish workers (Svedmyr, Melén and Kjellén, 1956) not only recovered adenoviruses from stools but also from mesenteric lymph nodes and suggested that the viruses may be associated with mesenteric lymphadenitis. In a study of 82 cases of such adenitis, however, only 6 were found to be excreting adenoviruses (Kjellén, Sterner and Svedmyr, 1957).

(*c*) *Atypical pneumonia and pneumonitis*

Another clinical variant is the suggestion that some cases of atypical pneumonia may be due to the adenovirus (Hilleman and Werner, 1954; Dascomb and Hilleman, 1956). Atypical pneumonia was thus recorded by Loffler and others (1956) in cases of acute respiratory disease among recruits in Switzerland. The radiological abnormalities pictured by these authors are similar to the aspiration pneumonia of Ramsay and Scadding (1939). The patients did not develop increased serum titres of red cell agglutinins demonstrable in the cold. Sporadic cases of so-called atypical pneumonia or pneumonitis in children and in adults associated with adenoviruses have also been recorded by workers in France (Lelong *et al.*, 1956), Sweden (Sterner, 1956), Italy (Pennacchio and Danese, 1956) and Britain (Pereira and Kelly, 1957b). In at least one of these patients the association of the pneumonic state with adenovirus infection depended upon the recovery of the virus at autopsy (Lépine *et al.*, 1956). In a more recent publication Chany and others (1958) have reported an outbreak of pneumonia in infants in a Public Assistance ward of a hospital in Paris. Eight deaths occurred and Type 7 adenovirus was incriminated.

(*d*) *Kerato-conjunctivitis*

The conjunctivitis due to adenovirus infection can occur independently of pharyngitis. Thus adult patients with kerato-conjunctivitis yielding adenovirus from eye swabs have been

of respiratory disease among University students and nurses (Tyrrell *et al.*, 1956) but there were 4 instances suggestive of this infection among 66 adults admitted to hospital in the winter of 1953–4 and suffering from acute exacerbations of bronchitis or of pneumonia. A recent longitudinal study conducted by Dr. Sutton in an infants' home in Sheffield during the winter of 1957–8 gave several strains of adenoviruses from throat swabs but the recoveries were achieved when the children were in their normal state of health. Most of these viruses belonged to Type 2 which seems to have a particular propensity to the latent state.

Combining the evidence of the role of adenoviruses in human disease it seems probable that different types vary in their behaviour. Types 3, 4 and 7 have been chiefly associated with outbreaks of respiratory disease in children and young adults whereas Types 1, 2 and 5 have usually been found in apparently healthy subjects. It is necessary to describe this phenomenon of virus carriage or latency in more detail.

Adenoviruses as Latent Agents

The earliest recovery of an adenovirus was effected from human adenoids, and according to Huebner, Bell and Rowe (1957) 90 per cent of surgically removed adenoids and tonsils yield adenoviruses by the technique of cultivation of tissue fragments. The viruses cannot be demonstrated in ground-up tissue presumably because of the presence of antibodies which are found in the sera from the same children. Six types of adenoviruses were found in the adenoids or tonsils by these workers but Types 1, 2 and 5 comprised 90 per cent of the strains. Occasionally, however, the viruses have been recovered under circumstances suggesting actual disease. Thus Pereira and Kelly (1957b) succeeded in recovering a Type 1 virus from the trachea and lung of a fatal case of pneumonitis in a 4½-month-old child. They also demonstrated the variant Type 7 strain from adenoids removed at Salisbury some months after an outbreak of pharyngo-conjunctival fever among children in the same town due to the same agent.

It therefore appears probable that the latent adenoviruses

the infection with these viruses cannot be associated with clinical disease at all. Thus in Cleveland only 10 of 531 pharyngeal swabs from children and adults with respiratory illness yielded adenoviruses, though increases in antibody titre against one or other of Types 1 to 7 occurred in from 1 to 5 per cent of cases during the same period (Jordan *et al.*, 1956). In the general population of three towns in Virginia and Maryland, Bell, Takemoto and Philip (Huebner *et al.*, 1957) studied 18,000 persons of whom 5,500 suffered from non-influenzal acute respiratory illnesses. Adenoviruses were isolated from 52 of 1,843 subjects. It was estimated that in children under 6 nearly 5 per cent of respiratory illnesses were attributable to adenoviruses. The figure for those aged 6 to 17 was 2·65 per cent and in adults it was only 0·3 per cent. This study lasted from September 1955 to April 1956, and it suggested that only a minor proportion of common febrile respiratory disease is due to the adenoviruses. It is particularly important to state that adenoviruses are not causally related to the common cold (Pereira and Kelly, 1957b).

The second method of study is that of conducting a longitudinal survey on the same persons over a period of time. Huebner, Bell and Rowe (1957) have conducted such a study in an orphanage in Washington where a good deal of change occurs in the population. Such an investigation is extremely time-consuming and involves the manipulation of enormous numbers of specimens. The authors found it difficult to relate the presence of adenoviruses in rectal or throat swabs with specific illnesses. They recovered viruses from 11 per cent of 2,066 throat swabs during a twelve-month period but only the Type 3 isolations appeared to be related to clinical events. It seems that the children were literally infested with adenoviruses most of which behaved as latent agents.

Studies in the general population in Britain have been meagre so far. In Glasgow, Grist and others (1957) found adenovirus infection only once in 70 cases of acute respiratory disease in a general practice and not at all in 103 adults and 32 babies admitted to hospital with pneumonia. In Sheffield no positive evidence of infection with adenoviruses was found in 40 cases

virus infection was demonstrated by Hilleman and his co-workers. Even the total febrile respiratory disease in the inoculated persons was considerably reduced, probably because adenovirus infection at Fort Dix maintains itself at a high endemic level. In the U.S. Navy a substantial reduction of febrile respiratory disease was also observed. Experiments are in progress in Britain with a similar inactivated monkey kidney vaccine and their outcome cannot be anticipated in this lecture. It seems to be agreed that no case can be made out for the use of such a vaccine in the prevention of common respiratory disease in the general population.

SUMMARY

It is easier to state the field of acute respiratory disease which is not apparently related to adenovirus infection than the converse. Adnoviruses are *not* the cause of the common cold or of influenza. They appear to be associated with outbreaks of febrile catarrh in Service communities and with febrile pharyngitis in children. Their role in the common respiratory disease of the community appears to be minor but there is evidence that sporadic respiratory disease in children may at times be caused by them. Certain strains of adenoviruses are commonly present in adenoids and tonsils removed for surgical reasons but their relationship to the chronic hyperplasia of these organs is obscure. It appears possible to prevent adenovirus infection in adults by inactivated tissue-culture vaccines.

REFERENCES

ANDREWES, C H, LAIDLAW, P. P and SMITH, W. (1935). *Brit. J. exp Path.* 16, 566.

ANDREWS, B E., MCDONALD, J. C., THORBURN, W. B and WILSON, J. S (1956) *Brit. med J.* 1, 1203

BALDUCCI, D, ZAIMAN, T E. and TYRRELL, D A J. (1956). *Brit. J. exp. Path* 37, 205

BARR, J, KJELLÉN, L and SVEDMYR, A (1958) *Acta paed.* 47, 365.

BELL, J. A., HANTOVER, M. J., HUEBNER, R. J. and LOOSLI, C G. (1956b). *J. Amer med. Ass* 161, 1521.

may be viruses persisting in host tissue months or years after an initial infection with the same agents. One may perhaps ask whether these latent agents are actually responsible for the hyperplasia of the adenoids or tonsils for which the organs were removed. This view has been suggested by some writers and also extended by Macfarlane and Sommerville (1957) to the lymphoid hyperplasia found in the lungs of children with bronchiectasis. Even if latent virus is not responsible for any pathological effects there is no doubt at all that such virus has retained its full virulence for man. Thus when introduced into the nose or swabbed on to the conjunctiva, viruses recovered from adenoids were as competent to produce clinical illnesses as other viruses recovered from patients with acute respiratory disease (Roden, Pereira and Chaproniere, 1956; Bell *et al.*, 1956a). These human volunteer experiments not only confirmed the fact that adenoviruses produce acute respiratory disease with or without conjunctivitis. They also showed that the possession of serum neutralizing antibodies in high titre before inoculation appeared to render the volunteers immune to infection accompanied by clinical illness.

A final warning concerning latent viruses is needed in that they may be recovered from specimens or tissues and falsely claimed to be the cause of actual illness. This is a danger common to other agents such as herpes virus.

Prevention of Adenovirus Infection

The elaboration of specific means of prevention of adenovirus infection may be premature at the present time when the role of these viruses in community disease is so obscure. Nevertheless, it appears that the adenoviruses of Types 3, 4 and 7 are regarded in the U.S.A. as an important cause of morbidity in the services. Trials have therefore been conducted in the Army (Hilleman *et al.*, 1957) and in the Navy by Bell and others (1956b). Both groups of workers have used formalin-inactivated virus vaccine prepared from monkey kidney tissue cultures. The titre of virus growth in this medium is not a high one, but the vaccine induced antibody production in those to whom it was administered. A very high degree of protection against adeno-

KENDALL, E. J. C., RIDDLE, R. W., TUCK, H. A., RODAN, K. S., ANDREWS, B. E. and McDONALD, J. C. (1957). *Brit. med J.* ii, 131.
KJELLÉN, L., LAGERMALM, G , SVEDMYR, A. and THORSSON, K.G. (1955). *Nature*, 175, 505.
KJELLÉN, L., STERNER, G. and SVEDMYR, A. (1957). *Acta paed.* 46, 164.
KJELLÉN, L , ZETTERBERG, B. and SVEDMYR, A. (1957). *Acta paed.* 46, 561.
LELONG, M., LÉPINE, P., ALISON, F., LE-TAN-VINH, SATGÉ, P. and CHANY, CH. (1956). *Arch. Franc. ped* 13, 1.
LÉPINE, P., CHANY, CH. and ROBBE-FOSSAT, F. (1956). *Ann. L'Inst. Pasteur*, 91, 607.
LÖFFLER, H , SPENGLER, G. A., RIVA, G , STUCKI, P. and MANGOLD, R. (1956). *Schweiz. med. Wschr.* 86, 967.
[illegible], W. W. and
[illegible] 770.
[illegible] HUANG, J. S (1957). *J. biophys. biochem. Cytol.* 3, 505.
MORGAN, C., HOWE, C., ROSE, H. M. and MOORE, D. H. (1956). *J. biophys. biochem. Cytol.* 2, 351.
MUNRO-ASHMAN, D , GARDNER, P. S , TAYLOR, C E. D. and McDONALD, J. C. (1958). *Lancet*, ii, 121.
OKER-BLOM, N., WAGER, O., STRANSTRÖM, H., MÄKELÄ, P. and JANSSON, E. (1957). *Ann. med. exp biol Fenn*, 35, 342.
PARROTT, R. H., ROWE, W. P., HUEBNER, R J , BERNTON, H. W. and [illegible]
[illegible] *Lancet*, ii, 592.
ROSEN, L (1958). *Virology*, 5, 574.
ROWE, W P , HARTLEY, J. W. and HUEBNER, R. J. (1958). *Proc. Soc. exp Biol N Y*, 97, 465.
ROWE, W P., HUEBNER, R. J , GILMORE, L. K , PARROTT, R. H. and WARD, T. G. (1953). *Proc. Soc. exp. Biol. N.Y* 84, 570.
ROWE, W P , SEAL, J. R., HUEBNER, R. J , WHITESIDE, J. E , WOOLRIDGE, R L and TURNER, H C. (1956). *Amer. J Hyg* 64, 211.
SOHIER, R., BENSIMON, P., CHARDONNET, Y., CHALLUT, F., and FREYDIER, J. (1957). *Rev Hyg. med. soc.* 5, 423.
SOMMERVILLE, R. G. (1957) *Proc. Roy Soc. Med.* 50, 757.
STERNER, G (1956) *Acta paed.* 45, 449.
STUART-HARRIS, C H. (1953). *Influenza and other virus disease of the respiratory tract* E Arnold, London.

BELL, J. A., ROWE, W. P., ENGLER, J. I., PARROTT, R. H. and HUEBNER, R. J. (1955) *J. Amer. med. Ass.* **157**, 1083

BELL, J. A., WARD, T. G., HUEBNER, R. J., ROWE, W. P., SUSKIND, G. and PAFFENBARGER, R. S. JR. (1956a). *Amer. J. pub. Hlth.* **46**, 1130

BENNETT, F. M., LAW, B. B., HAMILTON, W. and MACDONALD, A. (1957). *Lancet*, ii, 670.

BOYER, G. S., LEUCHTENBERGER, C. and GINSBERG, H. S. (1957). *J. exp. med.* **105**, 195.

CHANY, C., LÉPINE, L., LELONG, M., LE-TAN-VINH, SATGÉ, P. and VIRAT, J (1958). *Amer. J. Hyg.* **67**, 367.

COMMISSION ON ACUTE RESPIRATORY DISEASES
(1946). *Bull. Johns Hopkins Hosp.* **79**, 97.
(1947a). *Medicine*, **26**, 441.
(1947b). *J. Clin. Invest.* **26**, [illegible]

[illegible] M.R., [illegible] *d.* **72**, 367.

FISHER, T. N. and GINSBERG, H. S. (1957) *Proc. Soc. exp. Biol. N.Y.* **95**, 47.

FORSSELL, P., LAPINLEIMU, K., STRANDSTROM, H. and OKER-BLOM, N. (1956). *Ann. med. exp. biol. Fenn.* **34**, 287.

FOWLE, A. M. C., COCKERAM, A. and ORMSBY, H. L. (1955). *Amer. J. Ophthalmol.* **40**, 180.

GINSBERG, H. S., BADGER, G. F., DINGLE, J. H., JORDAN, W. S. JR. and KATZ, S. (1955). *J. clin. Invest.* **34**, 820.

GRIST, N. R., SOMMERVILLE, R. G. and CARSON, H. G. (1957) *Brit. med. J.* i, 378.

HILLEMAN, M. R., ANDERSON, S. A., LEVINSON, D. J. and LUECKING, M. L. (1958). *Amer. J. Hyg* **67**, 179

HILLEMAN, M. R., STALLONES, R. A., GAULD, R. L., WARFIELD, M. S. and ANDERSON, S. A. (1957). *Amer. J. Pub. Hlth* **47**, 841.

HILLEMAN, M. R. and WERNER, J. H. (1954) *Proc. Soc. exp. Biol. N.Y.* **85**, 183.

HUEBNER, R. J., BELL, J. A. and ROWE, W. P. (1957) 'Cellular biology, nucleic acids and viruses' *Publ. N.Y. Acad. Sci* **5**, 393.

JAWETZ, E., HANNA, L., KIMURA, S. J. and THYGESON, P. (1956). *Arch. Int. Med.* **98**, 71

JORDAN, W. S. JR., BADGER, G. F., CURTISS, C., DINGLE, J. H., GINSBERG, H. S. and GOLD, E. (1956) *Amer. J. Hyg* **64**, 336.

JORDAN, W. S. JR., BADGER, G. F. and DINGLE, J. H. (1958). *New Engl. J. Med.* **258**, 1041

KELLY, B. and PEREIRA, H. G. (1957). *Brit. J. exp. Path.* **38**, 396.

X

Epidemic Staphylococcal Infection in Hospitals[1]

R. E. O. WILLIAMS

THE current pandemic of interest in staphylococcal hospital infection has been responsible for many reports and memoranda on the importance of ascertaining the amount of staphylococcal disease in a hospital, and on administrative and technical methods of controlling it. These reports are mostly concerned with the 'endemic' level of infection, even though it has very often been an 'epidemic' of staphylococcal infection that has stimulated the interest. The epidemics have been submitted to surprisingly little analysis.

A search in the literature of the last twenty years has yielded a list (doubtless incomplete) of 53 records of epidemics, some very small and some very large. I have summarized these for the information that they give on the genesis of epidemic spread of staphylococcal infection. This analysis has been supplemented from the records of phage-typing carried out for hospital laboratories at the Staphylococcus Reference Laboratory, Colindale, during the four years 1954–7. Neither source of information is very satisfactory, the Colindale material because it often lacks epidemiological detail and the literature because of authors' natural preference for recounting detective stories in which the culprit is found, and a tendency to accept the simplest explanation, or the explanation that best fits with preconceived ideas The analysis seems, nevertheless, worth presenting because it offers some guidance as to the more profitable ways of

[1] The analyses summarized in this lecture have been presented in greater detail in a paper published in *The Lancet* (1959, i, 190).

STUART-HARRIS, C. H., ANDREWS, C. H. and SMITH, W. (1938). Med. Res Coun. Spec. Rep. Series no. 228.
SVEDMYR, A., MELÉN, B. and KJELLÉN, L. (1956). *Acta med. scand.* 154, Supp. 316, 20.
TANAKA, C. (1958). *Arch. Ophthalmol.* 59, 49.
[illegible]
WELLER, T. H., ROBBINS, F. C and ENDERS, J. F. (1949). *Proc. Soc. exp. Biol. N.Y.* 72, 153.
WENNER, H. A , BERAN, G. W., WESTON, J. and CHIN, T. D. Y. (1957). *J. inf. Dis.* 101, 275.
WOOLRIDGE, R. L , GRAYSTON, J. T., WHITESIDE, J. E., LOOSLI, C. G., FRIEDMAN, M. and PIERCE, W. E. (1956). *J. inf. Dis.* 99, 182.

In 27 of the outbreaks some search was made for the person who introduced the infection or who was responsible for its spread, and the authors thought they had found one such person in 13. This person, almost always a nurse, had a septic lesion in 5 cases, and was a healthy carrier in the other 8. In four of the epidemics the evidence implicating the carrier seems to have been good but in most of the others the attribution is only a reasonable probability; in only two was exclusion of the carrier the only and successful method of control adopted.

Baby-to-baby spread was presumably concerned in the 5 incidents in which no carrier could be implicated, while in the 7 with larger numbers of carriers two-way traffic in staphylococci seems very likely—babies infecting nurses, and nurses babies. In the last group of epidemics were those with the greatest numbers of babies affected.

Surgical wound infections

Fifteen accounts of epidemics of surgical wound infection have been found (Table 2); the smallest involved 6 patients and the

TABLE 2. Source of infection in 15 surgical ward and general outbreaks

Place	Carriers considered important	No. of outbreaks
Operating theatre	1 septic surgeon	1
" "	1 healthy surgeon	2
" "	1 or 2 healthy nurse(s)	2
Operating theatre + wards	1 healthy surgeon + septic patients	1
Ward	1 septic patient	1
Ward	(Several carriers importance uncertain)	2
Operating theatre + wards	" " " "	1*
Operating theatre	(Not sought)	1*
(Not sought)	" "	4

* Airborne infection considered important.

largest more than 230. Eleven had a single epidemic type of staphylococcus; one had 2 types and one many.

Six outbreaks were due to infection of patients, in the operating theatre, from one particular individual. In one a surgeon

investigating epidemics, and because it may perhaps stimulate the investigation and recording of a much greater number of epidemics, which is necessary for a full understanding of their origins.

For the present discussion no attempt has been made to define an 'epidemic'; the term has been used for those circumstances that prompted publication or investigation as examples of an undue prevalence of infection. This is tantamount to regarding as the distinction of epidemic from endemic prevalence of hospital infection the fact that the former is sufficient to draw attention to itself, while the latter is only revealed by specific inquiry. Such a distinction is very dependent on local circumstances and one man's epidemic must often be another's endemic level.

Published Records of Epidemics

Maternity units

Records of 32 epidemics have been discovered (Table 1); the smallest had 4 patients affected, the largest had at least 320.

Table 1 Source of infection in 32 maternity unit outbreaks

Carriers		No. of epidemics
Considered source of infection	Considered unimportant	
1 septic nurse	0–14	4
1 septic patient (?)	3	1
1 healthy nurse	0	4
1 healthy nurse	2–6	4
2–5 healthy nurses (?)	—	2
None	0	1
None	2–3	4
5–57 (importance uncertain)		7
Source not sought		5

In all cases the infection seemed to have spread in the ward or nursery, and all but one of the 32 had a single staphylococcus type responsible for all, or at least for the great majority of the cases.

staphylococci profusely on his forearms as well as on his hands and in his nose, and to some extent the frequency with which he punctured his gloves may have offered the special opportunity. But it seems very unlikely that either of these characteristics is rare and one is left wondering whether it was the staphylococcus that was exceptional. The carrier described by Shooter and others (1958) did not carry unusually profusely, but in this case the extent to which the staphylococcus spread among the patients in the ward pointed strongly to a high degree of that unexplained quality of 'communicability'. Hare and Ridley (1958) have ascribed potent dispersal in some persons to perineal carriage, but this characteristic has not been studied in any of the published work.

An extraordinary ability to spread among the staff was a characteristic of several outbreaks in both maternity and surgical units (Table 3). Why spread of this sort should sometimes occur

TABLE 3. Numbers of carriers of epidemic type in published epidemics

Type of epidemic	No. of epidemics with specified no. of carriers					Total epidemics
	0	1	2–3	4–9	10+	
Maternity units	1	6	8	7	5	27
Surgical units etc.	0	4	3	0	3	10

is quite unexplained, and certainly deserves further investigation. The complexity of these epidemics almost precludes discovery of the 'source', and one suspects that the spreaders of infection may be numerous and constantly reinforced as the initial victims infect further members of the staff. Although the outbreaks with numerous carriers have generally been the more extensive outbreaks, the ability of a strain to colonize the nose is not necessarily an indication that it is virulent, in the sense of producing disease, as was shown in our recent studies in a surgical ward (Shooter *et al.*, 1958).

The operating theatre is the one situation in which we have good evidence for airborne transfer of infection. The fact that airborne particles containing staphylococci, apparently liberated in the wards, can be carried on air currents to the operating

with a septic lesion on his hand was implicated, and 6 of 8 patients operated on by him in the relevant period developed sepsis. There are two other outbreaks attributed to surgeons who were healthy carriers and two to healthy carriers among the nursing staff. In one outbreak most of the cases were attributable to a healthy carrier in the surgical team, though other patients were certainly infected in the ward, probably from infected patients who were actively dispersing staphylococci.

Ward cross-infection was clearly important in two outbreaks, one in a urological unit and one in a general surgical unit. In the latter the infection was apparently introduced by a patient with a septic lesion, though not all the cases were infected directly from him. In neither of the other two outbreaks in which ward spread was considered probable, was it possible to recognize one individual as the source of the infections and there was presumably patient-to-patient infection.

The remaining two outbreaks in surgical wards implicated the air supplied to the theatre as the vehicle of infection of some or all of the patients.

The Genesis of Published Epidemics

The published records are certainly not a random sample of epidemics; it is, however, very striking how often a single individual was either responsible for all the cases in these outbreaks, or was responsible for introducing the infection to the unit and communicating it directly to some of the patients. When the individual is a surgeon or nurse and has a septic lesion it is not surprising that he or she infects the patients, but the examples of spread from healthy carriers—in several cases quite unequivocal—raise interesting queries. At least half the staff of any hospital carry staphylococci in their noses. One asks whether those who spread infection were carrying more extensively or in some other way differently from those who did not; or whether their work or behaviour afforded some unusual opportunity for spreading; or whether their staphylococcus was more 'virulent' or more 'communicable' than the others'. The healthy surgeon described by Devenish and Miles (1939) appeared to carry

stump must be nearly as susceptible to infection as an open wound, and far more difficult to protect.

The literature, then, indicates clearly that epidemics are often due to a single type of staphylococcus, and can, not infrequently, be attributed to a particular carrier—diseased or healthy. And one has the recurrent impression that the remarkable feature is not the frequency of staphylococcal disease but its rarity—unless we are being misled when we regard all staphylococci as of equal 'virulence'. Some further light can be thrown on these points by analysis of the staphylococci received for type-identification at Colindale.

Outbreaks due to Staphylococci Typed at Colindale

From the records of the four years 1954–7, we have available 200 separate incidents when a hospital pathologist sent us strains for typing because he or she thought there might be an epidemic of an infaection.

The total of 200 includes all incidents from which 3 or more strains from patients' lesions were sent for typing, with the exception of a few where we could not be certain which strains came from lesions and which from carriers, or where more than one hospital was involved in one incident and the different sources could not always be distinguished.

Of the 200 incidents, 53 had only 3 or 4 strains from lesions, but 46 had 20 or more. The total number of strains is not a very good guide to the real magnitude of an epidemic because it sometimes happened that strains could be recognized by their peculiar antibiotic sensitivity so that not all were sent to us; and on the other hand, some hospitals continued to send material even after the epidemic had subsided (and in so doing some encountered further examples of epidemic spread due to different types).

Strains were considered to belong to one type if their phage patterns did not differ by more than one 'strong' reaction (see Williams and Rippon, 1952; Williams, 1957a). The basic set of typing phages listed by Anderson and Williams (1956) was used throughout the four years, with the one exception that phage 80 was not brought into regular use until mid-1954.

theatre shows that they cannot be large, and the numbers of staphylococcus-containing particles in the air of an operating theatre can hardly ever be such that more than a very few of them would be expected to settle into the wound or on to its immediate surroundings during an operation. If such particles can infect the wound, the minimal infecting dose can hardly be more than a few hundred cocci and there are some indications (Lidwell *et al.*, 1959) that it is more likely to be in the tens. Although the surgeon-carrier could obviously insert large numbers of staphylococci into a wound it is more difficult to see how a nurse, even if assisting at the operation, could insert more than a few. But if very small numbers of cocci are in fact able to cause infections, it is again difficult to see why infections are not more common than they seem to be. Perhaps the relative rarity of operation infection again reflects a relative rarity of truly virulent staphylococci. This might seem unlikely when Shooter and others (1956) recorded no fewer than 11 different types in 17 wound infections. But St. Bartholomew's is a large hospital and the air to the theatre there was clearly heavily contaminated by air which could have come from several surgical wards in which there must have been many staphylococci whose virulence was attested by the sepsis that they had produced.

Only three of the published reports from surgical units give evidence for cross-infection within the wards—a phenomenon well demonstrated for the haemolytic streptococcus when it was the pathogen of the day (cf. Miles *et al.*, 1940). On the other hand all the maternity-unit outbreaks were regarded as ward-spread. It is well known that practically all infants born in hospital acquire staphylococci within their first week or so, and the indications are that at any one time one particular type is dominant (e.g. Hutchison and Bowman, 1957). It appears that there is ordinarily free transfer of staphylococci from one infant to another within the maternity unit; presumably an epidemic of disease occurs when the strain that is being transferred has greater invasive powers than the average. That the introduction, where traceable, is so often due to a nurse carrier is hardly surprising in view of the many contacts that the nurses have to have with the infants. The infant's nose, skin and umbilical

carriers than the others. It may be recalled that, even apart from epidemics, staphylococci of phage group I are the most common among nurses, and tend to be responsible for their septic lesions (Alder, Gillespie and Thompson, 1955).

TABLE 5. Carriers of epidemic type in Colindale epidemics (maternity and surgical)

Epidemic type of staphylococcus	No. of epidemics with specified no. of carriers 0	1	2–3	4–9	10+	Total epidemics
Group I, type 80	5	4	6	4	6	25
,, ,, others	5	2	5	4	3	19
Group II	5	2	2	0	1	10
Group III	14	11	9	1	0	35
Miscellaneous	3	0	2	0	0	5
Total	32	19	24	9	10	94

Frequency of phage types of staphylococci

Differences between the phage-types of staphylococci have been alluded to several times; it was therefore of interest to analyse the types of the 3,803 staphylococci isolated from cases of sepsis in hospital, and especially from epidemics of hospital sepsis, examined over the 4-year period. Almost all the strains from the maternity wards came from septic skin and eye lesions among babies, but some were from breast abscesses in the mothers. The strains from surgical and other units were mostly from wound sepsis, but some were from pneumonia, urinary and gastro-intestinal tract infections, and skin lesions.

Of the 3,803 strains 167 were untypable. When all the strains of one type isolated in a single epidemic were counted as one strain, we had available 1,131 'independent' strains for analysis. Altogether well over 200 different phage types were recognized among these 1,131 strains, but only 20 types or groups of closely related types occurred more than 10 times; the most common of these are shown in Table 6.

There is, first, a striking difference between the distribution of strains from maternity and surgical units: strains of phage group I were prominent in maternity units, and strains of phage group III in surgical units. The difference was especially

For the present analysis a type was regarded as an 'epidemic type' (E.T.) if it was isolated from three or more different patients' lesions.

Epidemic types

In about one third (34·5 per cent) of the incidents, typing failed to reveal any one type causing 3 or more cases of infections; 50·5 per cent of the incidents had one dominant epidemic type, and the remainder (15 per cent) had 2 or more (in 5 cases, 4 or more) such epidemic types. The proportion of mixed-type epidemics is thus much higher than is indicated by the literature. But no incident with 20 or more cases, and only about 30 per cent of the incidents with 5–19 cases, failed to show at least one E.T. (Table 4). In the incidents with 4 or more cases and

TABLE 4. Frequency of one or more epidemic types in Colindale material

	No. of lesion strains per incident		
	3–4	5–19	20+
No. of incidents	53	103	44
,, ,, ,, with 1 or more E.T.	16	71	44
Per cent of lesion strains in E.T.	96	57	66

at least one epidemic type, rather over 60 per cent of the lesion-strains examined belonged to the epidemic type; the balance was presumably evidence of 'endemic' infections forming a background to the epidemic.

Staff carriers

Cultures were received from carriers among the hospital staff from 94 of the epidemics that had one or more E.T. (Table 5). *In 32 no carrier of the E.T. was found;* 19 had one carrier and 24 had 2 or 3 carriers. Whether epidemiological studies confirmed these as the 'source' of the epidemic is not, in general, known.

Only 10 epidemics had 10 or more carriers. Nine of these outbreaks were due to staphylococci of phage group I; within phage group I, type-80 outbreaks had not, on average, more

results from a group of young adult males indicate that about 12 per cent of all the strains are untypable, compared with 4·5 per cent of the lesion strains, and that no more than 37 per [illegible]

[illegible]

tributed 32 per cent of the strains to Table 6.

Type 80

The existence of specific epidemic types has been exemplified especially well by 'type 80'. The specific phage for this strain was isolated in Australia in 1953 (see Rountree and Freeman, 1955) at a time when an extensive outbreak was occurring in a Sydney hospital. Soon afterwards many other outbreaks due to the same type were reported from hospitals throughout Australia, and it was noticed that the lesions due to type 80 were peculiarly severe. Shortly afterwards a very similar phage (81) was described in Canada (Bynoe, Elder and Comtois, 1956). Most type-80 strains seen in Britain are also lysed by phage 81.

Phage 80 was taken into the routine typing set at Colindale in the middle of 1954 and in that year the type was recognized in a total of 7 incidents, and infected 3 or more patients in 5 incidents. By 1957 we recognized the type in 39 incidents and as an epidemic type in 19—nearly half the incidents which had any recognizable epidemic type. No fewer than 30·1 per cent of all the lesion strains typed in 1957 proved to be type 80. And the increase has been maintained in 1958.

The spread of type 80 is not confined to Britain. Almost all the epidemics (mostly in maternity units) described in the United States in the last 2–3 years have been due to this type or others closely related to it (although commonly referred to in the literature as 52/42C/44A/80/81 etc.). Type 80 is common as an epidemic strain in the Netherlands and in eastern Europe and in Uganda. We are in fact at present observing the pandemic spread of type 80—a strain that has more recorded epidemics to its credit than any other type; whose epidemics are, in our experience, twice as extensive as the average of all other types; and which is especially prone to generate a high rate of

striking with the strains from epidemics: in fact half the maternity unit outbreaks were due to strains of group I, compared with a quarter of the surgical and other outbreaks. And indeed only one type within group I, namely type 80, produced any substantial number of outbreaks in surgical units.

TABLE 6. Phage type distribution of staphylococci from septic lesions

	Phage type	Maternity No. of *strains*	Maternity Per cent *epidemic*	Surgical & General No. of *strains*	Surgical & General Per cent *epidemic*
Group I	29	13	0	12	0
	79	20	25 0	17	0
	52A/79	42	38·1	34	11·8
	80	34	55 9	51	35·3
	Others	60	11·6	114	1 8
	Total	169	27·8	177	13 6
Group II	71	20	35 0	6	33·3
	Others	57	12·3	27	11·1
	Total	77	18 2	33	9·1
Group III	6/7/47/53/54/75 etc.	23	13 0	23	17·4
	7/47/53/54/75 etc.	29	24 1	45	13 2
	47/53/75/77	11	27·3	48	25 0
	75/77	2	(50)	37	48·2
	75	2	0	22	9·1
	53	15	6·7	30	13 3
	Others	87	9 2	138	11 6
	Total	169	13 6	343	18 0
Others		78	87	85	1·2
Total		493	18 2	638	14 0

Both in maternity units and in the surgical units there was a clear indication of the existence of epidemic types of staphylococci. Indeed 4 types (80, 52A/79, 71 and 7/47/53/54/75) were responsible for 50 per cent of the maternity unit outbreaks and 3 types (80, 75/77, 47/53/75/77) were responsible for 50 per cent of the outbreaks in the surgical wards. These groups of types were responsible for only 25 per cent or 21 per cent respectively of all the septic lesions. This difference is emphasized when the epidemic strains are compared with strains from healthy carriers. In collaboration with the Epidemiological Research Laboratory at Colindale we have recently examined the types of staphylococci harboured by healthy carriers. Preliminary

A few studies have been made of the characteristics of staphylococci isolated from lesions, compared with those isolated from carriers and other non-pathological situations (e.g. Schwabacher, Cunliffe, Williams and Harper, 1945; Hinton and Orr, 1957). These showed a statistical difference—the strains from pathological lesions were rather more active than those from carrier sites. But such studies inevitably suffer from the defect that we do not know whether, had the nasal strains gained access to wounds, they might not have produced sepsis.

Studies are needed to compare epidemic strains of staphylo-

[illegible]

a strain of type 47/53/75/77 is currently producing cases of sepsis in a hospital, it is to be regarded, at this time and in this place, as an epidemic type, and treated as such. But given the introduction of a single case of sepsis, or the discovery of a single carrier, we can only say that type 47/53/75/77 is on past experience more likely than, say, type 29 to produce epidemics. And the same is true of antibiotic-resistant strains.

Moreover only 25 per cent of the cases of type 47/53/75/77 sepsis recorded in Table 6 were associated with epidemic spread. The method of selection of the strains for this table makes it quite impossible to determine the real proportion of epidemic strains, but even allowing for this we may be sure that not all introductions, even of the most actively epidemic types, will be followed by spread of infection: the circumstances need to be propitious.

Prevention and Treatment of Epidemics

The review of published records of epidemics pointed strongly to the importance of particular 'dangerous' individuals in the genesis of outbreaks of infection. The analysis of the phage types endorses the view that, to explain the relative rarity of outbreaks of infection in relation to the abundance of carriers, the dangerous carrier must ordinarily be one who harbours a dangerous staphylococcus.

It might be that the only feature distinguishing the dangerous

carriage (and often also a high incidence of skin lesions) among the hospital staff.

Type 80 undoubtedly has something that many strains of staphylococci lack—but this fact should not be allowed to lead to the glib assumption that type 80 is the one and only epidemic staphylococcus. Other strains can often spread nearly as vigorously, and even in 1949 Colbeck reported an epidemic in maternity hospitals in Winnipeg due to type 57 in which the wide spread and, to some extent, the individual lesions closely resembled those due to type 80.

Characteristics of the epidemic types

It is inescapable that there are some strains of staphylococci distinguished by their ability to produce epidemics. The one feature that seems well established as characteristic of present-day epidemic strains is drug-resistance (see also Barber and Burston, 1955). Indeed it is difficult to see how a drug-sensitive strain could make much headway in the face of the amount of antibiotic likely to be used to control an epidemic. Most of the strains that are sent to Colindale for typing are in any case penicillin-resistant, but it is worth noting that 19 of 22 independent strains belonging to one or other of the three types (80, 47/53/75/77, 75/77) causing outbreaks in surgical wards were resistant to tetracycline as well as penicillin. Two of the types producing epidemics in maternity wards (52A/79 and 71) were rather rarely tetracycline-resistant.

We do not know (and it is difficult to see how we can ever discover) whether the antibiotic-resistant staphylococci that now spread through our hospitals are inherently more 'virulent' than their sensitive predecessors, but it is certainly possible that widespread drug treatment has tended to select drug-resistant variants of the more virulent strains, since it is the virulent staphylococci in patients with septic lesions that are most likely to be exposed to drug treatment. But clearly resistance to several antibiotics is not a prerequisite for an epidemic strain: type 80 when first seen in Australia was resistant only to penicillin, and epidemics of staphylococcal infections were observed before the discovery of penicillin.

the investigations that should be made (see also Williams, 1957b). The first must be to determine the types of the staphylococci from infected patients, to find out whether a single type is dominant. Epidemiological analysis may, with the help of the phage typing, suggest where the patients were infected—in operating theatre, ward or nursery. And then a search among staff (and patients) may detect persons who are carrying or infected with the epidemic type, one of whom might be the dangerous carrier who was responsible. Such a search will not elucidate all epidemics but it may well solve the problem in those in which management is the easiest (and which are disproportionately common in the literature)—those where all or most of the infection stems directly from one dangerous carrier.

More often, one may be sure, epidemics are not investigated until more than one source of infection is present, until the epidemic strain is being spread from one patient to another, or from several carriers to the patients. Then I believe that the guiding principle should be to treat the staphylococcal infections due to the specific epidemic type as one would treat cases and carriers of the agent of any other infectious disease—to segregate them from the community of susceptible patients until they are free of the specific infection, and to try to eliminate the epidemic type from the inanimate reservoirs in the hospital. Such a policy often strains both the physical resources of the hospital, to provide the necessary isolation facilities, and the laboratory services, to provide the necessary rapid recognition of the specific staphylococcus. But staphylococcal epidemics can greatly disrupt the work of the hospital, and it is very important that attempts should be made to control them, if necessary by the provision of extra facilities within the hospital or the use of those available outside. And it is to be hoped that accounts will be published of the attempts—whether successful or not—so that all may profit and that a future analysis of the literature may present a more representative view than is now possible of the epidemics as they occur.

carrier is the sort of staphylococcus he carries. The apparently especial danger of people with septic lesions might just reflect the fact that the lesion in the carrier is a built-in indicator of the virulence of his staphylococcus. We know of many instances in which carriers of notorious epidemic types fail to start epidemics—and in which epidemics cease even when carriers of the epidemic type are still present. Quite apart from the possibility that no susceptible patients are exposed, this could mean that there are variations in the virulence of the strains within the type; or it could be (and we must hope often is) that the aseptic precautions are sufficient to prevent spread even of the dangerous strains from the heavy disperser. But it could also be, as already suggested, that some people are dangerous because they disperse their staphylococci more widely than others. Probably all these explanations are appropriate to some cases, certainly the known differences in aptitude for dispersal by carriers (Hare and Ridley, 1958; Shooter *et al.*, 1958) are sufficient to make one feel that these differences must sometimes be relevant.

A hospital's routine precautions have to be designed on the assumption that any person may be a dangerous carrier and any staphylococcus a virulent one. If the precautions prove unable to prevent the emergence of epidemics—and this inability proves not to be simply due to defective applications of well-known precautions—we need to inquire whether there are more specific methods by which we could prevent epidemics. Can we recognize the dangerous staphylococci or dangerous dispersers of them? At present the answer must be: only, it seems, to a very limited extent. We know some of the sorts of people who have proved dangerous—but not how often such people are dangerous. We know which phage types of staphylococci have caused the most epidemics—but so many different strains have given rise to some epidemics that to confine attention to the notorious types would almost certainly be ineffective. We need a great deal more research into the biological characteristics of the staphylococcus that lead to virulence; and into the physiological (and behavioural) characteristics of carriers that make them dispersers.

But in the face of an epidemic, the analysis suggests some of

XI

The Regulation of the Blood Vessels in the Limbs[1]

A. D. M. GREENFIELD

ALL blood vessels exhibit a resistance to the flow of blood, and a capacity to hold blood which depends on the distending pressure. These two functions are not equally developed in the various categories of blood vessels. On the classical view resistance resides mainly in the arterioles, and capacity mainly in the venules and veins. But, of course, all of the vessels in a limb contribute to the resistance to the flow of blood through it, and the recent animal work of Kelly and Visscher (1956) and of Haddy, Fleishman and Emanuel (1957), which has been extended to man by Wallace and Stead (1957) and Patterson and Whelan (1955), emphasizes that, under at least some conditions, the normal arterial and venous trees may contribute quite largely to resistance to flow.

In what follows, we shall confine our attention to the behaviour of total limb resistance. There is not, at present, sufficient information to analyse the contributions of the various sorts of vessels to this total resistance. Nor, from the point of view of tissue nutrition or total body haemodynamics is such an analysis a matter of first importance. The two aspects of limb resistance with which we shall deal are firstly the response of resistance vessels to variations in distending or transmural pressure, and secondly the present state of our knowledge of the innervation and reflex control of these vessels. I would like to stress that in this presentation I am drawing on the contributions of many

[1] Two recent reviews (Greenfield, 1957; Roddie and Shepherd, 1958) deal with the subject-matter of parts of the content of this lecture

REFERENCES

ALDER, V. G., GILLESPIE, W. A. and THOMPSON, MARGARET E. M. (1955). *J. Path. Bact.* 70, 503.

ANDERSON, E. S. and WILLIAMS, R. E. O. (1956). *J. clin. Path.* 9, 94.

BARBER, MARY and BURSTON, J. (1955). *Lancet*, ii, 578.

BYNOE, E. T., ELDER, R. H. and COMTOIS, R. D. (1956). *Canad. J. Microbiol.* 2, 346.

[illegible]

HINTON, N. A. and ORR, J. H. (1957). *J. lab. clin. Med.* 50, 901.

HUTCHISON, J. G. P. and BOWMAN, W. D (1957). *Acta paed.* 46, 125.

LIDWELL, O. M., NOBLE, W. C. and DOLPHIN, G. W. (1959). *J. Hyg, Camb.* 57, 299.

MILES, A. A., SCHWABACHER, HERTA, CUNLIFFE, A. C., PATERSON ROSS, J., SPOONER, E. T. C., PILCHER, R. S. and WRIGHT, JOYCE (1940). *Brit. med. J.* Dec. 21 & 28, 855, 895.

ROUNTREE, PHYLLIS M. and FREEMAN, BARBARA M. (1955). *Med. J. Australia*, ii, 157.

SCHWABACHER, HERTA, CUNLIFFE, A. C., WILLIAMS, R. E. O. and HARPER, G J. (1945). *Brit. J. exper. Path.* 26, 124.

SHOOTER, R. A., SMITH, M A., GRIFFITHS, J. D., BROWN, MARY E. A, WILLIAMS, R. E. O, RIPPON, JOAN E. and JEVONS, M. PATRICIA (1958). *Brit. med J* i, 607.

SHOOTER, R. A., TAYLOR, G. W., ELLIS, G. and ROSS, SIR J. P. (1956). [illegible]

Assoc.

WILLIAMS, R. E O. and RIPPON, JOAN E. (1952). *J. Hyg., Camb.* 50, 320.

mechanisms have been suggested. The first is the venivasomotor reflex. It was suggested by Gaskell and Burton (1953) that a rise in pressure in the veins brings about a contraction in the resistance vessels. The second is a local and possibly myogenic response of the resistance vessels, causing them to narrow when the distending pressure is raised, and to widen when it is lowered. Such a mechanism has been proposed by Bayliss (1902) and Folkow (1949). There is a theoretical difficulty if it is assumed that the detector and effector are in the same cell; such a system might oppose but could hardly reverse the passive change in dimension. But this is not an insuperable obstacle. It is possible, for example, that the stimulus is a longitudinal stretching of the vessels, and the response a circumferential contraction. Even in the case of a passive vessel, distension may eventually be limited by inelastic elements in the walls.

Considering the critical closing pressure, we may say that the evidence on the whole supports the hypothesis that resistance vessels constrict completely at very low transmural pressures. Such low pressures are unusual in healthy people, and normally occur only at the extremities of raised limbs. Plethysmographically Gaskell and Burton (1953) found that the flow ceased in the toes on elevating the legs to 45° or 60° although the calculated arterial pressure available for perfusion was 47/17 mm. Hg. In agreement with this, Roddie and Shepherd (1957a) found that blood flow measured calorimetrically ceases when a finger is raised above the head until the mean perfusion pressure is lowered to between 35 and 60 mm. Hg. Reducing the vascular transmural pressure of a horizontal limb by local application of pressure above atmospheric, Yamada (1954) with the plethysmograph, and Roddie and Shepherd (1957a) by calorimetry have confirmed that flow ceases when the perfusion pressure is still about 40 mm. Hg. The critical closing pressure varies with the state of vasomotor tone, being greater when the tone is high, and lower when the tone is low. It might be expected that the vessels which had snapped shut would open again only in the face of a pressure much greater than that which allowed them to close. This is not so; the opening and closing pressures are almost identical.

laboratories, and that the contributions from Queen's University are in the main the work of my former or present colleagues, Professors Shepherd and Whelan, and Drs. Roddie, Patterson, Coles, Kidd and others.

The Response to Variations in Transmural Pressure

We will first consider vascular reactions to changes in the distending force, or, more precisely, to changes in transmural pressure (Fig. 1).

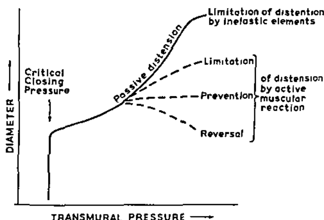


Fig. 1. The diameter of the lumen of a resistance vessel at various transmural pressures. For explanation, see text

Burton (1951) pointed out that, if certain reasonable assumptions are made about the properties of the muscular and elastic elements in the walls of small vessels, instability is to be expected below a certain transmural pressure. As the transmural pressure is reduced, it may be predicted that a point should be reached at which the equilibrium between the distending force and the tension in the walls of small vessels becomes unstable, allowing the vessels to close completely. The threshold pressure below which the vessels are closed Burton called the 'Critical closing pressure'.

As the transmural pressure is raised, resistance vessels would be passively distended unless there were an active reaction in their walls to limit, prevent, or reverse the distension. Two such

method almost certainly underestimates the true flow during even mild congestion, the unchanged flow with reduced perfusion pressure indicates a lowered resistance, and some dilatation of the resistance vessels. With more severe venous back pressure (Shanks, 1955; England and Johnston, 1956) calorimetric measurements indicate that the blood flow decreases as the perfusion pressure decreases, and that resistance and therefore calibre remain substantially unchanged.

There is, however, a good deal of evidence that vessels do in fact react actively to transmural pressure. The reaction is displayed when the increased transmural pressure is suddenly removed. The effect is rather like suddenly removing one team from a tug-of-war. The resistance vessels suddenly contract. Thus a period of fairly severe venous congestion, to 110 mm. Hg. for 5 minutes, sufficient to reduce the blood flow greatly, is followed not by a reactive hyperaemia, but by a period of reduced flow, indicating vasoconstriction. This is the case not only in normal, but also in sympathectomized and denervated arms (Patterson and Shepherd, 1954).

There is a similar response after a temporary increase in the transmural pressure in all the vessels of a part brought about by local exposure of the limb to subatmospheric pressures. The transmural pressure is increased by the amount by which the local pressure falls below atmospheric. Figure 2 shows the reduction in blood flow through the forearm immediately after exposure to −100 mm. Hg. for varying periods (Greenfield and Patterson, 1954a). Similar results are seen in the calf of the leg (Coles, Kidd and Patterson, 1956).

A further example of the after effects of altered transmural pressure is provided by the effects of arresting the circulation. When the circulation is arrested for a time and then released, a great increase in blood flow, known as reactive hyperaemia, follows. A contributory factor to this appears to be the fall in pressure in the arteries during the period of arrest. If, just before arterial occlusion, the arm is congested with blood by application

Let us turn now to increases in transmural pressure. A rise in transmural pressure is a very common happening. The pressure is more or less raised in all the vessels, arteries, capillaries and veins, below heart level. If this provoked a contraction of the local resistance vessels, whether by a venivasomotor reflex or a local myogenic response, we should have a mechanism in almost constant operation in some or other part of the body, and possibly playing an important part in the defence of the whole circulation against gravity. Since in a dependent inactive limb the arterial and venous pressures are about equally increased by the hydrostatic effect of columns of blood in the vessels extending to heart level, the perfusion pressure is essentially unchanged. Therefore, if the resistance vessels constrict the flow will be less than in a horizontal limb, and if they remain unchanged the flow also will be unchanged. Observations by venous occlusion plethysmography (Gaskell and Burton, 1953; Beaconsfield and Ginsburg, 1955a) have indeed indicated a reduction in flow, and therefore a contraction of the resistance vessels, in dependent limbs. The validity of this method when a part is already congested with blood is, however, doubtful, and, by contrast, other methods agree in indicating a small *increase* in the blood flow, and therefore dilatation of the resistance vessels, in the dependent limb.

For example, the heat elimination from the toes of a dependent leg is greater than that from the toes of the opposite horizontal leg (England and Johnston, 1956) and the same is true of the fingers (R. A. Roddie, 1955). Again, the oxygen saturation of both superficial and deep venous blood is increased when the arm is moved to the dependent position (Rosensweig, 1955) and this is also the case in the legs (Wilkins, Halperin and Litter, 1950). Evidently the reaction to increased transmural pressure does not prevent some widening of the resistance vessels in the semi-dependent limb, but it may oppose and limit this widening.

One reaches a similar conclusion when the effects of venous congestion are considered. Slight venous congestion, of up to 20 mm. Hg., which is sufficient to cause an increase in limb volume of about 2 per cent, does not cause any decrease in the

blood during local exposure to subatmospheric pressure indicate a progressive fall in flow, and therefore a narrowing of the vessels, in both the skin and muscle of the forearm (Blair and Roddie, 1958).

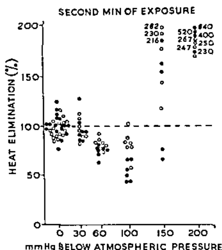


FIG. 3 The heat elimination from one hand during the second 5 minutes

1956.)

The most valuable information about a prolonged rise of transmural pressure comes from patients with coarctation of the aorta. The perfusion pressure is raised in the upper limb, but the blood flow is normal in both forearm and calf. The resistance to flow is increased in the forearm (exposed to the raised transmural pressure) but normal in the calf (Patterson, Shep-

normal pressure by the post-coarcted aorta. But it is improbable that a nervous or humoral mechanism for raising resistance would distinguish between these two beds. It is not certain that

of local subatmospheric pressure (Patterson 1956) or venous congestion (Wood, Litter and Wilkins, 1955), so that the arterial pressure is better sustained during the occlusion, the reactive hyperaemia is much reduced.

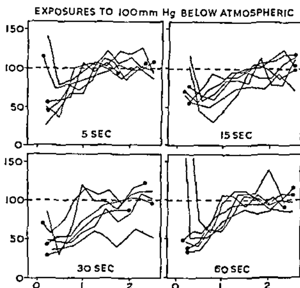


FIG. 2 The blood flow through one forearm following local exposure to a pressure 100 mm. Hg below atmospheric for various periods expressed (ordinate) as a percentage of the simultaneously measured blood flow through the opposite forearm at atmospheric pressure throughout. *Abscissa:* time in minutes from the end of the exposure. (From Greenfield and Patterson, 1954a.)

The next observations relate to the behaviour of the vessels while transmural pressure is increased. In acute experiments, blood flow has been followed calorimetrically in the hand (Fig. 3, Coles and Greenfield, 1956) and toes (Coles, 1957) during local exposure to subatmospheric pressure. These observations all indicate that when transmural pressure is increased beyond the amount normally found in full dependency, the resistance is increased. The results also indicate that in the fully dependent toes passive distension is successfully opposed, and the resistance is unchanged. *Observations on the oxygen saturation of venous*

and cholinergic nerves during body heating is excluded by the finding that atropine does not change the blood flow. The maintenance of an undiminished blood flow after nerve block is not due to persistence of a stable vasomotor material released before the block, for Roddie, Shepherd and Whelan (1957c) found that blocking the nerves before general body heating was started had no effect on heat elimination from the finger-tip.

Because of the enormous background flow through arteriovenous anastomoses, a modest vasodilator action on other vessels might well escape notice by the methods currently available, but we can feel sure that it plays no significant part in the temperature-regulating reactions of the normal circulation. However, since there are numerous sweat glands in the hands, the blood flow to which is probably increased in activity, it is possible that the nerves to the sweat glands of the hand have an indirect vasodilator action (Barcroft, Brod, Hayes and Hirsjarvi, 1959).

Apart from responses to changes in central body temperature, and to reflexes arising from temperature receptors in the skin (Cooper and Kerslake, 1953), the hand circulation is reflexly affected by stimuli from many regions of the body, by receptors activated on taking a single deep breath (Gilliatt, 1948) and by the emotions, but, strangely, it is practically excluded from participation in baroreceptor reflexes (Beaconsfield and Ginsburg, 1955b).

The skin of the forearm differs from that of the hand in being subject to vasoconstrictor nervous control, operating through a small range, and active while the subject is cold, and to vasodilator nervous control, operating through a large range, and active while the subject is hot. It is curious and remarkable that there is, seemingly, an abrupt change of control at the wrist.

Evidence for vasodilator nerves to forearm skin was first provided by Grant and Holling (1938) who found that in warm persons, when individual cutaneous nerves are blocked, the temperature of the skin is lower in the anaesthetized area than in the surrounding areas. Their existence has been confirmed by Edholm, Fox and Macpherson (1957) and by Roddie, Shepherd and Whelan (1957b). The latter found that the superficial

the increased resistance is a consequence of the increased pressure, but this hypothesis fits the facts better than the others that have been proposed.

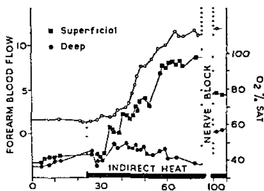
There is, therefore, a considerable body of evidence that resistance vessels react actively to an increase in transmural pressure on the arterial side of the circulation. This reaction must contribute to the protection of the circulation against the effects of gravity. To what extent it operates in, or contributes to or even causes arterial hypertension remains to be evaluated.

The Innervation and Reflex Control of the Resistance Vessels

Let us now turn to the innervation of the vessels. We shall refer exclusively to the upper limb; it is supposed that the conclusions here apply also to the lower limb, but this has not yet been adequately tested.

The skin of the hand, and especially of the digits, has been known for a long time to be under the influence of powerful vasoconstrictor nerves. The blood flow through the finger tips can be decreased to about one-hundredth part when these nerves are brought from quiescence to a state of full activity. This is probably the greatest range of blood flow encountered anywhere in the body. It has been suggested from time to time that there are also vasodilator nerves to the skin of the hand, but these are unimportant. The hand and finger circulation is, of course, increased by general body heating (Gibbon and Landis, 1932; Lewis and Pickering, 1931) and it has been presumed that the condition under which vasodilator nerves, if they exist, are most likely to be active, is during intense body heating. Evidence for their action would be obtained if, during such heating, nerve block reduced the blood flow through the blocked area, or, supposing the fibres to be cholinergic, if atropine did the same. Most careful observations have failed to disclose any vasodilator activity under these conditions in normal persons. Arnott and Macfie (1948) found no reduction in heat elimination from the nerve-blocked 5th finger, and Gaskell (1956) no reduction in blood flow through the blocked hand. The remote possibility of a balanced action of adrenergic

atropinization which effectively suppresses sweating. The exact relationship of the vascular response to sweat gland activity is not clear; it is partly linked to activity in the glands, but whether the so-called vasodilator fibres act directly on the vessels, or indirectly as a consequence of sweat-gland activity, or by both routes, is not yet known.



In muscle, the changes in blood flow that can be brought about by variations in nervous control are quite small compared with the changes due to exercise. Exercise of the muscles may raise the total forearm flow from 3 to 30 ml./100 ml./min. Barcroft, Bonnar, Edholm and Effron (1943) found that blocking the deep nerves raised the total forearm flow only to about 7 ml./100 ml./min. Of this increase a small part was in the skin, about one-third of which was anaesthetized, but most was in the muscles. The increase in the muscle flow was not a mechanical consequence of muscular paralysis, for blocking the deep nerves in sympathectomized subjects was without effect on flow.

venous blood attained a higher oxygen saturation (96 per cent) in the intact forearm than in the forearm with the superficial nerves blocked (85 per cent). The fairly small difference in venous oxygen saturation implies a very large difference in flow, perhaps in the ratio 5 : 1 or more. The presence of vasoconstrictor fibres to forearm skin is shown by the rise in oxygen saturation in the superficial venous blood when the nerves are blocked.

Understanding of the innervation of the forearm skin has been greatly assisted and simplified by the recent demonstration that the muscle circulation does not participate in the reflex adjustments to general body heating. This was shown to be so by Barcroft, Bock, Hensel and Kitchin (1955) who measured muscle blood flow with a calorimeter-sound, by Edholm, Fox and Macpherson (1956) who suppressed circulatory changes in the skin by intense adrenaline electrophoresis, and by Roddie, Shepherd and Whelan (1957b) who found that body heating raised the oxygen content of venous blood draining skin but not that of venous blood draining muscle (Fig. 4). All these observations agree with the findings of McGirr (1952) that Na_{24} clearance from muscle is unchanged, but that from skin is increased.

During body heating we may say therefore that blood flow through the muscle remains constant, and variations in total forearm blood flow indicate variations in the blood flow through the skin of the forearm.

When a rather cold person is heated (Roddie, Shepherd and Whelan, 1957a), *the forearm blood flow first increases to about* 5 ml./100 ml./min. This increase coincides with the increase of hand blood flow due to withdrawal of constrictor tone, is unaffected by atropine and is about equal to the increase seen on blocking the superficial nerves. The increase from 2 to 5 ml./100 ml./min. may therefore be taken to be due to release of cutaneous vasoconstrictor tone.

Continued heating leads to sweating in the forearm, indicated by a fall in skin resistance, and coinciding with this there is a further rise in total forearm flow from about 5 to 15 ml./100 ml./min. This increase is probably due to vasodilator nerves. The increase is reduced and delayed, but not abolished, by local

of the deep nerves abolishes the response, so the mechanism is clearly not humoral but nervous, and therefore operates by adjusting vasoconstrictor tone. In agreement with this conclusion, the blood flow on raising the legs approaches, but does not exceed, that in the opposite forearm with the deep nerves blocked.

The increase in blood flow is shown to be in muscle and not in skin by observations on the oxygen saturation of venous blood. The oxygen saturation of blood from the skin remains unchanged, but that of blood from the muscles is increased. Thus the response is brought about by a reduction in the vasoconstrictor tone of the muscular circulation.

The increase in muscle circulation is probably a reflex response to stimulation of baroreceptors in the low-pressure vessels in the chest. The evidence is as follows. Raising the cuffed legs is without effect until the cuffs are released. Thus the action depends on transfer of blood from the legs to the remainder of the body. Raising the legs and pelvis is more effective than raising the legs alone. Thus the [illegible]

[illegible] below the neck.

During leg raising the central venous pressure is raised, and so is the central venous pulse pressure, but the arterial pressure is hardly altered (Fig. 5). Thus the effect can hardly be a reflex from the arterial baroreceptors, and in any case, as we have mentioned, the muscle circulation is exempted from participation in baroreceptor reflexes from the carotid sinus.

A similar increase in muscle blood flow, due to a reduction of vasoconstrictor tone, is also brought about by resisted breathing (Roddie, Shepherd and Whelan, 1958). The subject breathes through a narrow orifice which causes the pressure in the air passages to fluctuate between about +30 and −20 mm. Hg. The effect of this is to set up cyclical strains in the tissues between the air passages and the blood vessels, and therefore most probably at baroreceptors situated in the walls of low-pressure intrathoracic vessels. This experiment suggests that the baroreceptors may be sensitive not only to changes in mean pressure,

Thus there is convincing evidence that there are nerves to the blood vessels of skeletal muscles, and that these normally exert a moderate degree of vasoconstrictor tone. This is confirmed by the increase in [illegible] blood (Roddie, Shepherd [illegible] deep nerves of the forear[illegible] exerting an intense vasoconstrictor action, for inhalation of 30 per cent carbon dioxide almost stops the blood flow through the normal forearm, although that through the nerve-blocked forearm is not greatly changed (McArdle, Roddie, Shepherd and Whelan, 1957). Since there is evidence from the oxygen saturation of venous blood draining the skin of the forearm, and from measurements of hand blood flow that the skin vessels are not intensely constricted, the constriction of the muscle vessels must be very acute indeed.

We have already mentioned the evidence that the vasoconstrictor tone to the muscle vessels is unaffected during general body heating. The muscle circulation is, of course, not well adapted for exchange of heat with the environment, and it is apparently excused from the circulatory adjustments of temperature regulation. More surprisingly, the muscle circulation is also excused from participation in reflexes arising from arterial baroreceptors in the carotid sinuses. Carotid compression which lowers both the mean and the pulse pressures at the carotid sinus does not increase the resistance to blood flow through the calf (Roddie and Shepherd, 1957b) and conversely, increasing the transmural pressure at the carotid sinuses by local external suction (Ernsting and Parry, 1957) probably does not reduce the resistance to blood flow through the hand, forearm or calf of the leg.

The vasoconstrictor tone to muscle is, however, readily adjusted in response to stimulation of baroreceptors in the low-pressure vessels in the chest (Roddie, Shepherd and Whelan, 1957d). Raising the legs of a normal recumbent subject (Fig. 5) increases the blood flow through the arm. Since the arterial pressure is scarcely changed, the increase must be due to vasodilatation. Atropinization is without effect, so cholinergic vasodilator nerves are not involved. Sympathectomy or block

nervously mediated changes in forearm blood flow. They may be responsible for the reduction in flow (Brigden, Howarth and Sharpey-Schafer, 1950) on assuming the erect posture, and for the reduction in flow after the Valsalva manœuvre, and the rise in flow after coughing described by Sharpey-Schafer (1953). The part played by the reflexes in congestive heart failure needs further evaluation; it may be that the receptors are set at an altered level.

None of the evidence we have considered indicates the existence of vasodilator nerves to the muscle blood vessels. There is convincing evidence in animals (Uvnas, 1954) for the existence of cholinergic vasodilator fibres to muscle, distributed through the sympathetic system, and discharged by electrical stimulation of various parts of the brain. In the human being the evidence for vasodilator fibres to muscle vessels rests on the observation that during fainting the total forearm blood flow is, for a short time, greater in the normal than in the nerve-blocked forearm (Barcroft and Edholm, 1945). If it may be assumed that the blood flow through the skin of the forearm is not increased in fainting, this evidence indicates a discharge of muscle vasodilators. It is not yet known whether more common circumstances than fainting cause the muscle vasodilators to discharge; this is a matter which we are at present investigating.

SUMMARY

The blood vessels of the skin of the hand are under very powerful vasoconstrictor control, and those of the skin of the forearm are under weak vasoconstrictor and powerful vasodilator control. In both areas of skin, the vessels are employed for thermoregulatory reflexes, but they are exempted from participation in baroreceptor reflexes.

The blood vessels of the muscles of the forearm are under vasoconstrictor control, and there is normally a state of vasoconstrictor tone which may be increased and almost stop the circulation, or relaxed, allowing the blood flow to increase to several times the normal value. This control is independent of thermoregulatory reflexes and of carotid sinus reflexes, but is employed by baroreceptor reflexes from the low-pressure vessels

but also to the extent of rhythmical pressure fluctuations. Since, however, the site of the receptors is unknown, we do not know exactly how much of the intrathoracic pressure disturbances are transmitted to them.

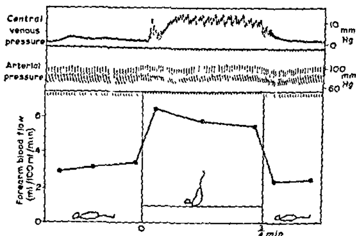


Fig 5. The effect of raising the legs and lower trunk on the central venous pressure, the arterial pressure and the forearm blood flow (from Roddie, Shepherd and Whelan, 1957d).

In the recumbent subject the reflex mechanism is set at a level which allows the muscle blood flow to be either increased or decreased. This is shown by experiments in which the subject breathes from a mouthpiece delivering air at a steady pressure *above or below atmospheric (Blair, Glover and Kidd, 1958).* From the point of view of low-pressure baroreceptors, reducing the pressure in the air passages is presumably equivalent to raising the pressure in the great veins. Both procedures raise the forearm blood flow. Conversely, however, raising the pressure in the air passages reduces the forearm blood flow. Breathing against a raised pressure may be presumed to have the same effect on low-pressure baroreceptors as rising from the horizontal to the erect posture.

Reflexes from the low-pressure baroreceptors, rather than from arterial baroreceptors, may be responsible for a number of

ENGLAND, R. M. and JOHNSTON, J. G. McC. (1956). *Clin. Sci.* **15**, 587.
ERNSTING, J and PARRY, D. J (1957). *J. Physiol.* **137**, 45P.
FOLKOW, B (1949). *Acta physiol. scand.* **17**, 289.
[illegible]

GRANT, R. T. and HOLLING, H E. (1938). *Clin Sci* **3**, 273.
GREENFIELD, A D. M. (1957). *Amer. J Med.* **23**, 675.
GREENFIELD, A. D. M and PATTERSON, G. C. (1954a). *J. Physiol.* **125**, 508.
GREENFIELD, A D. M. and PATTERSON, G. C. (1954b) *J. Physiol.* **125**, 525.
HADDY, F. J , FLEISHMAN, M. and EMANUEL, D. A. (1957). *Circulation Res.* **5**, 247
KELLY, W. D and VISSCHER, M B (1956). *Amer. J. Physiol.* **185**, 453.
LEWIS, T. and PICKERING, G. W. (1931). *Heart*, **16**, 33
McARDLE, L , RODDIE, I. C., SHEPHERD, J T. and WHELAN, R F. (1957). *Brit. J. Pharmacol.* **12**, 293.
McGIRR, E. M. (1952). *Clin. Sci.* **11**, 91.
PATTERSON, G. C. (1956) *Clin. Sci.* **15**, 17.
PATTERSON, G C. and SHEPHERD, J. T. (1954). *J. Physiol.* **125**, 501.
PATTERSON, G. C., SHEPHERD, J T. and WHELAN, R. F. (1957). *Clin. Sci.* **16**, 627.
[illegible]
[illegible]
[illegible]
[illegible]
[illegible] *J. Physiol.* **136**, 489.
RODDIE, I C., SHEPHERD, J T. and WHELAN, R. F. (1957b). *Clin. Sci* **16**, 67.
RODDIE, I C , SHEPHERD, J. T. and WHELAN, R. F. (1957c). *J. Physiol* **138**, 445.
RODDIE, I C , SHEPHERD, J T. and WHELAN, R. F. (1957d). *J Physiol.* **139**, 369
RODDIE, I. C., SHEPHERD, J T and WHELAN, R. F. (1958) *Circulation Res* **6**, 232.
RODDIE, R. A. (1955). *J. appl. Physiol.* **8**, 67.
ROSENSWEIG, J. (1955). *J. Physiol.* **129**, 281.
SHANKS, R. G (1955). *Clin. Sci* **14**, 285.
SHARPEY-SCHAFER, E. P. (1953). *J. Physiol.* **122**, 351
[illegible]

WOOD, J. E., LITTER, J and WILKINS, R. W. (1955). *Circulation Res.* **3**, 581.
YAMADA, S (1954) *J. appl. Physiol.* **6**, 495

in the chest. Vasodilator nerves to the forearm have so far only been shown to act during fainting.

Note

Since this lecture was given, Blair, Glover, Greenfield and Roddie (*J. Physiol.* 1959, **148**, 633) have found that emotional stress activates cholinergic vasodilator nerves to muscle capable of causing a very great increase in blood flow, of the same order as that seen after exercise.

Acknowledgements

The author's thanks is due to the authors, and to the editors, of the *Journal of Physiology* and *Clinical Science* for permission to reproduce illustrations.

REFERENCES

Arnott, W. M. and Macfie, J. (1948). *J. Physiol.* **107**, 233.
Barcroft, H., Bock, K. D., Hensel, H. and Kitchin, A. H. (1955). *Pflug. Arch. ges. Physiol.* **261**, 199.
Barcroft, H., Bonnar, W., Edholm, O. G. and Effron, A. S (1943). *J. Physiol* **102**, 21.
Barcroft, H., Brod, J., Hayes, J. P. L. A and Hirsjarvi, E. (1959). *J. Physiol.* **145**, 5P
Barcroft, H. and Edholm, O G. (1945). *J. Physiol.* **104**, 161.
Bayliss, W. M (1902). *J. Physiol* **28**, 220.
Beaconsfield, P. and Ginsburg, J. (1955a) *Circulation Res.* **3**, 478
Beaconsfield, P. and Ginsburg, J. (1955b) *J Physiol.* **130**, 467.
Blair, D. A., Glover, W. E. and Kidd, B. S. L (1958). *J Physiol.* **140**, 40P.
Blair, D A and Roddie, I C. (1958) *J Physiol.* **143**, 67P.
Brigden, W., Howarth, S and Sharpey-Schafer, E. P. (1950). *Clin. Sci* **9**, 79.
Burton, A. C. (1951). *Amer. J. Physiol.* **164**, 319.
Coles, D. R. (1957). *J. Physiol.* **135**, 171.
Coles, D. R. and Greenfield, A. D. M. (1956). *J. Physiol.* **131**, 277.
Coles, D R. and Kidd, B. S L. (1957). *Circulation Res.* **5**, 223.
Coles, D R., Kidd, B S. L. and Patterson, G. C. (1956) *J. Physiol* **134**. 665
Cooper, K. E. and Kerslake, D McK (1953). *J. Physiol.* **119**, 18.
Edholm, O. G., Fox, R H. and Macpherson, R. K. (1956). *J. Physiol.* **134**, 612.
Edholm, O. G., Fox, R. H and Macpherson, R. K. (1957). *J. Physiol.* **139**, 455.

decrease in the inflammatory response is associated with a decreased resistance to infection. A boxer, on the other hand, has no teleological scruples about calling it a bad thing. The anti-phlogistic effect of a raw steak clapped to his bruised cheek, followed by judicious incisions to relieve the oedema, means that he has an eye to see out of for his next day's fight. If you object to folk-lore from the prize-fighter's dressing-room as evidence that inflammation is a curable evil, there are in orthodox medicine plenty of more deeply considered—though perhaps equally folk-lorish—instances of anti-phlogistic therapy with a fair claim to success.

Even the most sceptical will agree that the processes of repair at the finish of the inflammatory sequence can be considered as biologically useful; and even the most teleological will agree that the injury which initiates the sequence is biologically useless; but in between there is a complex of interwoven events whose use to an organism is difficult or impossible to determine. The *a priori* principle, that all the events we observe are purposeful, even when they are not successful, is very little help in thinking about inflammation, because the only objective criterion of purpose is achievement, and it is obvious that many inflammatory processes end in disaster.

As a beginning then, let us cautiously admit that inflammation may be good in parts. But in this context of biological usefulness we must scan the inflammatory sequence with a cold eye and be prepared to dismiss some events as mainly neutral, to castigate others as nothing but a nuisance, and perhaps to find a hard core of reactions that appear to be a necessity for the ultimate recovery of the inflamed tissue. Many of the attempts to unravel inflammation are bedevilled by the acceptance, sometimes unconscious, of the notion that all parts of the phenomenon are relevant. The chief pitfall of the notion is not that it is wrong, but that it uses time and energy that might

[illegible]

selves with this question—'Is it neutral, a nuisance or a necessity?' The answer can determine whether a given event is necessary if inflammation as such is to occur; and whether what

XII

Mediators of the Vascular Phenomena of Inflammation

A. A. MILES

If I begin by declaring this lecture to be an illustration of the theme, 'the scientific basis of medicine', it is not because I am succumbing to that all-too-common desire of lecturers to tell their audiences things they already know quite well; but because I wish to emphasize a point of view—namely, that I should attempt to be scientific not only about the mediators of inflammation, but about medicine as well. The aim of medicine being to preserve or restore health, our analysis of inflammation should be made with an eye on what it implies for the preservation or restoration of health. In saying this I wish to raise no false hopes of arriving at conclusions of supreme clinical importance. I make the point because the sort of eye we keep on general medicine will depend in the first place on what we feel about inflammation. Is it a good or a bad thing?

The Biological Value of Inflammation

The confirmed teleologist, for whom all tissue responses are purposive, will maintain it is a good and necessary part of the process leading to recovery from injury. An immunologist might say it is good because, under the stimulus of an infecting microbe, the permeability of the blood vessels increases to let through into the extravascular tissue the proteins and phagocytes capable of killing the invader. He would accordingly deprecate the use of cortisone for anti-phlogistic purposes, at least in an animal like the rabbit, because he has proved that the

becomes still more permeable to plasma, and permits the passage of white blood cells, while the stickiness is exaggerated into the formation of intravascular adhesive matter which entrains platelets and the larger blood cells. After injury of this degree, [illegible]

result, there may be rupture of the wall, and haemorrhage. How useful is this extravascular change? A freer movement of intercellular fluid might be necessary for the inflammatory sequence, but the haemorrhage is clearly an irreversible change and a nuisance and can be dismissed from our scheme.

There is little more to be said about the possible benefits of this 'unit' extravascular change. Hyaluronidase is the only respectable candidate for mediator (Zweifach and Chambers, 1950); and there are, in fact, indications that endogenous hyaluronidase accumulates in inflamed tissues (Mayer, 1950). [illegible]

standing extravascular response—the degranulation or rupture of the basophilic mast cells present in a wide variety of tissues, and its pharmacological consequences—is dealt with below.

Vascular Events in Early Inflammation

There is much more to be said about the vascular events, and particularly about the capillary change. Its place in the time sequence established by microscopic observation is important. When the micro-injury is inflicted away from the vessels the liquidation follows after a time-lag, and capillary permeability usually increases after that, strongly suggesting mediation by diffusible substances formed in response to the injury—formed, that is, *extrinsically* with respect to the vascular endothelium. Another important feature of the sequence, borne out by experiments on a macroscopic scale, is that increased permeability and stickiness of the capillaries go hand in hand (e.g. Burke and Miles, 1958), and appear to be necessary preludes to the diapedesis of leucocytes.

we finally decide to be the essential inflammation is necessary for recovery after injury.

To simplify the issue, first let us dismiss all the processes of cellular repair and concentrate on the earlier events proper to inflammation. Second, let us confine ourselves to inflammation without tissue-destruction—what is sometimes called 'physiological inflammation' because recovery takes place without formation of new tissue. It is possible that qualitatively different events take place when injured tissues proceed to necrosis, but since reversible degrees of inflammation can nearly always be produced by weaker versions of necrotizing stimuli, we should not go far wrong with the lesser phenomenon.

The 'Unit' Events of Inflammation

From this phenomenon we can select as it were structural and functional units of which the typical inflammation is built. The anatomical unit appears to be a microscopic region of tissue, the capillary that feeds it, and the appropriate arterioles and venules supplying and draining the capillary. The immediate response to micro-injury of a unit of this kind, as seen in the transparent tissues of a tadpole's tail (Clark and Clark, 1935) or the rat meso-appendix (Zweifach, 1953) is in two parts—an extravascular and a vascular. These are the two functional units and they are distinguished as such for the following reasons. A few minutes after injury, applied in areas away from the blood vessels, the connective tissue, hitherto a resistant gel through which fluids may creep in a parsimonious way, apparently along minute perifibrillar spaces, becomes so fluid that injected material can move freely for relatively large distances. Very soon after this the nearby arterioles dilate, and the capillary walls—especially in the venous region—become sticky for circulating leucocytes, and become more permeable to plasma proteins. The extravascular 'liquidation' (if I may steal an apt word from politics) is independent of the vascular change in the sense that it is not, for example, induced by histamine, whereas the capillary events are; and because it can be mediated by injected hyaluronidase, which does not increase capillary permeability. With greater injurious stimuli the capillary wall

Nevertheless, permeability to proteins usually increases at an early stage in the vascular response, so that it is a useful index of at least an initial capillary reaction, if not of a continuing one. None of these various indicators is perfect, and their various imperfections must necessarily influence our interpretation of permeability effects.

Capillary Structure and Permeability Changes

Just what do we expect a mediator of increased permeability to do? The plasma protein has to pass at an increased rate either through the endothelial cell or the gaps between adjacent endothelial cells, and then through the pericapillary layer of connective tissue. In some situations this layer is fibrous and in others appears to be no more than intercellular ground substance—a substance that only in situations like the lung or the kidney, where the capillaries lie next to the epithelium, achieves the status of a definite membrane. The passage between cells is apparently blocked by material, revealed by staining with silver, which has been called 'cement substance'. In electronmicrographs of capillaries, however, there is no sign of a special cement. In sections of lung capillaries, for example (Karrer, 1956), the alveolar and endothelial cells lie on each side of the basement membrane, and the very narrow space between endothelial cells appears to be filled with a substance continuous with that of the basement membrane. In other capillaries the interendothelial substance is a continuation of the pericapillary sheath. When histamine is applied to the vessels in doses big enough to induce a permeability effect, the flat endothelial cells round up and project into the lumen (Zweifach, 1953). A similar rounding-up is seen in electron-micrographs, presumably a consequence of the fixation of the tissue for section. How this structural change after histamine could contribute to increased permeability is not clear, because, even with the much greater insult of fixation, a thin sheet of endothelial cytoplasm still lines the wall of the vessel with minimal gaps between cells. We may suppose the endothelium to be impermeable and permeability an increase in the gaps between cells, whether achieved by changes in the porosity of the connective tissue

For these reasons, increased capillary permeability, which of all the events in this unit response is one of the easiest to measure dynamically in the living animal, is popular as an indicator reaction in inflammation and in the search for mediators of inflammation.

The detection of increased permeability

Its precise measurement, whereby a balance sheet is established for blood supply, accumulation of blood substances in the tissues, and the contents of the lymph draining the affected part, demands laborious techniques which are rarely applied. For the most part, an adequate and reasonably constant drainage may be assumed in acute experiments, and the accumulation in the tissues either of fluid, or of excessive amounts of large-molecular substances, is used to indicate increased permeability. The available indicator reactions vary in subtlety and range. One of the simplest, and still useful, was first explicitly described by Mr. Tony Weller, on the occasion of his forced attendance at a tea-party of the Brick Hill Lane Branch of the United Grand Junction Ebenezer Temperance Association: 'There's a young 'ooman,' he said, 'as has drunk nine breakfast cups and a half; and she's swelling wisibly before my wery eyes.' The modern version of Mr. Weller's indicator reaction is the rapid swelling of the rat's snout or paw under the influence of permeability factors given either locally or systemically. The commonest method is the accumulation of dye in the skin of the test animal at the site of injury. The dye is given intravenously and in the blood becomes closely associated with plasma proteins, particularly the albumins (Rawson, 1943). Rapid extravasation of dye is therefore accepted as an index of rapid exudation of protein. When a more firmly marked protein is required, albumin or globulin can be tagged, for example, with radioactive iodine. In other methods the injury is applied to a serous cavity—pleural or peritoneal, and the resulting exudates harvested for examination.

As indicators, the plasma proteins are misleading in some ways: thus, capillaries may be relatively impermeable to the proteins but pathologically permeable to other substances.

Nevertheless, permeability to proteins usually increases at an early stage in the vascular response, so that it is a useful index of at least an initial capillary reaction, if not of a continuing one. None of these various indicators is perfect, and their various imperfections must necessarily influence our interpretation of permeability effects.

CAPILLARY STRUCTURE AND PERMEABILITY CHANGES

Just what do we expect a mediator of increased permeability to do? The plasma protein has to pass at an increased rate either through the endothelial cell or the gaps between adjacent endothelial cells, and then through the pericapillary layer of connective tissue. In some situations this layer is fibrous and in others appears to be no more than intercellular ground substance—a substance that only in situations like the lung or the kidney, where the capillaries lie next to the epithelium, achieves the status of a definite membrane. The passage between cells is apparently blocked by material, revealed by staining with silver, which has been called 'cement substance'. In electronmicrographs of capillaries, however, there is no sign of a special cement. In sections of lung capillaries, for example (Karrer, 1956), the alveolar and endothelial cells lie on each side of the basement membrane, and the very narrow space between endothelial cells appears to be filled with a substance continuous with that of the basement membrane. In other capillaries the interendothelial substance is a continuation of the pericapillary sheath. When histamine is applied to the vessels in doses big enough to induce a permeability effect, the flat endothelial cells round up and project into the lumen (Zweifach, 1953). A similar rounding-up is seen in electron-micrographs, presumably a consequence of the fixation of the tissue for section. How this structural change after histamine could contribute to increased permeability is not clear, because, even with the much greater insult of fixation, a thin sheet of endothelial cytoplasm still lines the wall of the vessel with minimal gaps between cells. We may suppose the endothelium to be impermeable and permeability an increase in the gaps between cells, whether achieved by changes in the porosity of the connective tissue

barrier, or an uncovering of more barrier with a fixed porosity. J. R. Pappenheimer (1953) calculated that intercellular gaps provided a sufficient surface for the normal passage of solutes, proteins and other colloids, and postulated that filtration took place through an isoporous membrane. If this is so, increased permeability would be equivalent to increase in the number of available pores. It is tempting to think of the outraged endothelial cell drawing in its skirts, and thereby exposing, at the intercellular junctions, more of an underlying filtration surface; but our only evidence that these endothelial sheets may be disturbed in this way is the supposed passage of phagocytes *between* endothelial cells. Other work (Shirley *et al.*, 1957) suggests, however, that the size of these pores can be increased—a process that might, for example, be brought about by depolymerization of the barrier substance—whatever that may be.

These notions about the mechanism of filtration have the merit of not requiring a change in the endothelial cell—the events could take place solely in the intercellular gaps. Other speculators lean to increased activity of the endothelium. Some electron-microscopists describe pores in the endothelial sheet (Pease, 1955; Rhodin, 1955), others depressions that are not quite complete holes (Karrer, 1956). By themselves, however, if we accept their dimensions as reflecting what they are like in real life, the holes are probably too big to be responsible for selective filtration of proteins; though they might well be the site of grosser leaks of plasma. The presence of 'vacuoles' in endothelial cytoplasm and of cups on the cell surface suggesting the active intake of blood fluid by the process of pinocytosis, and the demonstration that visible particles that have stuck to the surface are ingested (albeit very slowly compared with the processes we are considering) give some, but not much, plausibility to the notion of transfer of blood substances to the extravascular tissues by pinocytosis. But, of course, the transfer of proteins, whether at a physiological or a pathological rate, might well occur on a submicroscopic scale as part of a metabolic process, and independently of a relatively coarse process like pinocytosis.

My object, however, is not to weigh probabilities about

mechanisms but to indicate that the mediators we seek may be anything from a depolymerizing enzyme to a small-molecular substance triggering a profound change in the transporting mechanisms of the endothelial cell.

Candidates for the Role of Endogenous Mediators of Increased Permeability

The more notorious substances of endogenous origin that increase permeability are listed in Table 1. Although histamine

Table 1 Some candidates for the role of endogenous permeability factors

1. (*a*) Pharmacologically active amines:	
	Histamine
	5-Hydroxytryptamine (5-HT) and
(*b*) Their liberators	
2. (*a*) Proteases:	
	Plasmin
	Serum globulin PF
(*b*) Products of proteolysis:	
	Leukotaxine
	Bradykinin
	Kallidin

was recognized by Eppinger as a permeability factor as early as 1913, interest in its possible role in inflammation was stimulated principally by Lewis's observation in 1927 that the triple response, elicited by pricking histamine into human skin, was also induced by mechanical, electrical, thermal and chemical injury. Histamine is regarded as the active end-product of a series of reactions which either liberate histamine from a chemical bondage or which set free already-formed histamine from special sites in a cell. Serotonin—or 5-hydroxytryptamine—is a newcomer, which was in 1956 shown conclusively by Rowley and Benditt to be a powerful permeability factor in the rat, an animal in which, incidentally, histamine has a low permeability-increasing potency. The notion of proteases, especially plasma proteases, as mediators of permeability effects was widely entertained during investigations of the anaphylatoxin theory of anaphylaxis (Friedberger, 1909; Jobling and Peterson, 1914).

Proteases appear in the blood during antigen-antibody combination *in vivo* and can be induced *in vitro* by adding to fresh serum substances like agar, starch and bacterial suspensions. Proteases could act directly on the capillaries, but they must also be considered as precursors of polypeptide mediators. From the early nineties onwards many investigators induced permeability effects with products of tissue breakdown. One type of permeability factor—leukotaxine—was found in such tissues and was described by Menkin in the 1930's (see Menkin, 1956). He found it in one-day-old alkaline inflammatory exudates, and identified it as a polypeptide. It appears that Menkin's leukotaxine is representative of a family of polypeptides (Duthie and Chain, 1939; Spector, 1951), having in common a moderate permeability-increasing potency, dependent mainly, as Spector showed, upon the number of amino-acid residues in the molecule—some 12–15 being optimal. Both kallidin, a polypeptide released from plasma by the pancreatic protease kallikrein, and bradykinin, a polypeptide released from plasma by treatment with trypsin or snake venom, increase capillary permeability in the experimental animal.

Under proteases, I should mention two more of the numerous mediators of inflammation named by Menkin—exudin and necrosin. Being largely [illegible] fractions of inflammato[illegible] terized for detailed cons[illegible]ation. Nevertheless, we should note that necrosin, from mature inflamed tissue, may be a tryptic enzyme (Gorkin, 1957), and Menkin's exudin effects may well have been due to the globulin permeability factors discussed in the next paragraph. Plasmin—that is, fibrinolysin—is included in the list because, being formed from its precursor plasminogen during treatment of plasma to produce anaphylatoxins, it has been put forward as a mediator of inflammation. But in our hands it has proved to be neither a strong permeability factor on its own, nor is it a protease unequivocally proved to liberate histamine or other permeability factors.

The Serum Proteases Increasing Capillary Permeability

The serum permeability proteases, on the other hand, have a relatively high permeability-increasing potency in their own right, and as good a *prima facie* claim as histamine to the role of natural mediators of permeability effects. Our knowledge of them, which is far from complete, is based largely on work in three laboratories—by Paskhina (1956) in Moscow, by ourselves at the Lister Institute (Miles and Wilhelm, 1955, 1958; Wilhelm *et al.*, 1955, 1958; Wilhelm, Mill and Miles, 1957; Elder and Wilhelm, 1958; Mill *et al.*, 1958; Becker, Wilhelm and Miles, 1959), and by Spector (1956, 1957) at University College Hospital (see also Stewart and Bliss, 1957). Most of this work was done in the guinea-pig, rat, rabbit and man, but there are good reasons for believing that it holds also for cattle, horses, swine, dog, cat and mouse—that these permeability proteases are, in fact, a characteristic feature of all mammalian plasma.

Normal plasma contains an inert pro-factor (symbolized by pro-PF), which in most species is readily activated to PF by dilution of fresh serum in glass tubes, or by exposing neat serum to finely divided glass (Margolis, 1958), to cellulose, to starch granules, or to agar.

(The activable serum globulin permeability factor, by the way, has not been named; to avoid this circumlocution, it is referred to as 'serum PF' or 'protease PF'.)

The list of PF activators is, of course, reminiscent of those activating anaphylatoxins, but we need not overestimate the coincidence, because other apparently distinct factors—the 'pain-producing substance' of Keele (Armstrong *et al*, 1957), and the 'contact' factors concerned in blood coagulation (Margolis, 1957)—are similarly activated. Our pro-PF indeed is but one of a number of blood substances activated in this way. The important feature is the ready activation of pro-PF to form PF.

The PFS, and presumably the pro-PFS, are either α- or β-globulins, according to the species of animal, and some of them, even in the relatively impure state of our best preparation, are equal to histamine in potency, although their molecular size is 2,000 times as great. This suggested to us that they were

enzymes, and all our subsequent work has confirmed the suggestion. We have not yet found an *in vitro* substrate nor identified the substrate they attack in the body, but they are inhibited by a wide variety of trypsin-inhibitors from plant seeds and by the esterase poison diisopropylfluorophosphonate in concentrations similar to those which inhibit trypsin. Also we have indications that they attack certain artificial amino-acid esters (Becker *et al.*, 1959). For these reasons we talk of the serum PFS as proteases and believe they attack linkages similar to those attacked by tryptic enzymes. Nevertheless, they have no frank proteolytic activity, which suggests that their attack is restricted to particular linkages in the protein molecule—in this respect resembling, for example, the attack of thrombin on fibrinogen.

To complete the picture, we should note that the serum of each of the species investigated contains a slowly acting inhibitor of PF (symbolized by IPF).

The Relation of Endogenous Permeability Factors to One Another

A gallant attempt to bring many of these candidates into a consistent scheme was made by Ungar (1956; Fig. 1). At I, the three kinases symbolize the stimulus of, respectively, the damaged cell, the anaphylactic serum, or infecting bacteria. The reality of stages II and III is based on the action of certain inhibitors. Stage III clearly accords well with our notions about the pro-PF, the activated PF protease and the native inhibitor (IPF) present in serum. But there the resemblance ceases, because we have no evidence that the PF proteases yield an active polypeptide and none that they liberate histamine. This last negative result accords with results of other work on histamine releasers, namely, that histamine does not appear to be liberated directly from a combination with protein. One could add to the diagram by inserting an arrow from peptides to histamine, because many polypeptides are in fact histamine-liberators.

Since Ungar's scheme was conceived the work of Riley and West (1953) has established the tissue mast cell as one of the most abundant and ready sources of not only histamine but also

of 5-hydroxytryptamine; it is now clear that both can be liberated by all kinds of direct stimuli, ranging from distilled water to lecithin-breaking enzymes, which degranulate or rupture these basophilic cells, without, apparently, the intervention of any proteolytic mechanism.

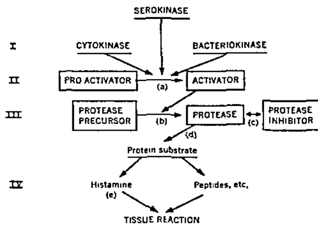


FIG 1. Diagram of the biochemical sequence leading to the tissue reaction associated with histamine release. Letters (a) to (e) indicate points at which the reaction can be blocked. (After Ungar, 1956.)

Most of the evidence at present points to the independence of mediation by proteolysis and its consequences, and mediation by liberators of histamine and 5-hydroxytryptamine.

CRITERIA FOR ESTABLISHING A PERMEABILITY FACTOR AS A NATURAL MEDIATOR

We have, then, three main classes of candidates for the role of mediator—histamine and 5-HT, the protease PF globulins of plasma (which do not appear to release active polypeptides), and the permeability polypeptides. To establish any one of them as a natural mediator, we have two kinds of evidence at our disposal (Table 2). The first kind consists of facts without which the notion of mediation would not even be plausible, and the second, of actual proofs of mediation.

Universality of species distribution may not be relevant if we wish to solve the problem for a given species of animal, which might have its own peculiar mechanism of response. But it is important if we seek, as we do, to establish a common mechanism. There is little difficulty with any of the candidates. Histamine, 5-HT and the protease PF are found in a wide variety of animals, and we may expect polypeptides wherever there are proteins.

TABLE 2. Criteria for the implication of an endogenous permeability factor in inflammation of vertebrate tissue

Contributions to plausibility	
(1) Universality of species distribution	
(2) Potency	
(3) Availability in tissues	
	(*a*) Quantity
	(*b*) Activability
(4) Availability of natural inhibitors	
(5) Mimicry of inflammatory response	
Contributions to proof	
(1) Presence in tissue *at the relevant time*	
(2) Effect of depletion	
(3) Effect of specific inhibitors	

Potency is sometimes crucial. It is generally expected, on the assumption that nature is moderately economical, that physiologically or pathologically effective substances will be highly potent and, therefore, that of a number of possible factors, the most potent is likely to be the naturally effective one. This is not necessarily true, but if a mediator has a low potency, it must at least be available in amounts commensurate with the observed natural effects. The figures in Table 3 record the permeability effect of the candidates, measured as the number of standard permeability-inducing doses per milligram of preparation, as tested in three species of laboratory animal. The serum PFs are highly or reasonably potent, especially in the animal of origin. So is histamine, but at this stage 5-HT becomes a casualty, at least as a common mediator, because in amounts likely to be liberated in the tissues it is active only in the rat (Sparrow and Wilhelm, 1957), though in this animal, whose capillaries are relatively insusceptible to histamine, it may well play the role

that histamine plays in the guinea-pig and rabbit. The polypeptides, at any rate in the crude preparations available to us, have a low potency.

TABLE 3. Potency, measured as the number of 'Effective Blueing Doses' (EBD)/mg, of various endogenous permeability factors in the skin of three laboratory animals

Permeability factor	'Effective Blueing Doses' per mg. when tested in		
	guinea-pig	rat	rabbit
Guinea-pig globulin PF	38,700	1,100	130
Rat ,, PF	1,400	620	400
Rabbit ,, PF	930	120	22,000
Trypsin	2,600	900	600
Kallikrein*	240	70	1,130
Leukotaxine †	65	nt	nt
'Leucotaxine'‡	320	120	80
Bradykinin§	170	50	300
Histamine	32,200	1,400	37,000
5-Hydroxytryptamine	60	16,200	20

nt=not tested

* from pig pancreas
† from pleural exudate in dog
‡ from a peptic digest of fibrin
§ from guinea-pig plasma protein

Availability in tissues is considered under two heads, the amount of precursor and the readiness with which it is activated by injury. There is plenty of histamine in the tissue and plenty of protein as precursors for polypeptides. There is also plenty of activable pro-PF in the serum (Table 4). However, except as a reservoir from which tissue pro-PF and PF might be derived, the large amount in the blood is irrelevant to mediation that *starts* in the extravascular tissues. We have investigated this point only in the guinea-pig, but there it is clear that both the pro-PF and the inhibitor IPF are present in the intercellular tissue spaces and, in fact, share with all the plasma proteins the usual property of passing in small amounts from the normal capillary through the tissues to the lymph vessels. We estimate that the pro-PF in normal skin would yield 10 times more protease PF than is needed to affect all the capillaries of a given region.

The extravascular situation of liberable histamine and 5-HT is well established. A great deal of it is concentrated in the paravascular mast cells and the relation of readily damaged mast cell, release of histamine and the vascular response to it, is regarded by some as a neat device of Nature for a microhormonal response of small blood vessels to local injury. It *is* neat, but we must not be impressed too much by the coincidences. It may be the mast cell's misfortune, not its privilege, to be so incontinent with its histamine; and vertebrate anatomy being what it is, proximity to blood vessels is a situation that nearly all the cells of the body, including the mast cell, would find hard to avoid.

TABLE 4 The content of activable protease permeability factor (PF) in one millilitre of the serum of various mammals

Source of PF	Activable PF 'Effective Blueing Doses' per ml.
Man*	2,300
Guinea-pig	30,000
Rat	3,000
Rabbit	20,000

* Tested in the guinea-pig, the remaining factors tested in the homologous animal.

As regards activability, the mast cell is easily persuaded to degranulation and rupture, but not as easily as is commonly assumed. Synthetic histamine-liberators are usually strongly basic or highly surface-active substances—like octylamine and Tween 80—and other liberators, like distilled water or the α-toxin of *Cl. welchii*, are often even more biologically brutal; suggesting that liberation involves a greater energy exchange than, for example, the surface activation of a pro-PF by glass. Moreover, although degranulated mast cells are found in sections of inflamed tissue, inflammation can occur without their general degranulation. Smith and Miles (1960) studied the early stages of bacterial infection in the peritoneal cavity of the rat, which is lined with abundant mast cells packed with histamine and 5-HT. As you will see from Table 5, both Locke's

solution and a suspension of dead staphylococci in Locke's solution induced exudation of protein increasing up to the fourth hour; and living staphylococci a greater exudation, with moderate diapedesis of leucocytes. But in all situations and at the times when tests were made not more than 10 per cent of the histamine available in the mast cells was liberated, and in spite of active infection, there was no sign of their degranulation or rupture. They were equally resistant to staphylococcal α-toxin, and to infections with *C. ovis.*

TABLE 5. The response of the rat peritoneal cavity to bacterial injury

Stimulus	Protein content* and (percentage histamine released) in rat peritoneal exudates after		
	½ hr.	3 hr.	4 hr.
Locke's solution	—	—	3·1 (9%)
Staph aureus. dead	2·7 (6%)	5·0 (9%)	4·8 (8%)
Staph aureus. living	2 9 (8%)	6·4 (11%)	5 8 (11%)
Staph. aureus toxin	—	—	12 4 (*c.* 8%)
Staph. aureus toxin + antitoxin	—	—	4 0 (*c.* 8%)
C ovis. living	—	—	9 0
C. ovis. living (antihistamine)	—	—	10 8
Locke's solution (antihistamine)	—	—	6 2

* Expressed as absorptiometer readings.

I have stressed that permeability protease is readily activable by the contact of the pro-PF with a foreign surface, and it is plausible to suppose that tissue cells injured only to the point of a change in surface properties, without gross internal derangements of metabolism, could activate profactor in the intercellular fluid. I once cast the polymorphonuclear leucocytes for this role, but unfortunately they could not be induced to play the part.

The rapid production of a polypeptide presents no difficulty. At first sight, the degradation of protein to polypeptide might seem a slow process of activation for damaged tissues to depend on. There is no biochemical reason, however, why this should

not be rapid—and no good reason why the polypeptide precursor should not be a simpler compound than a full protein. Hilton and his colleagues have provided a physiological model for the rapid production of a polypeptide in their discovery that the stimulation of salivary (Hilton and Lewis, 1956) and sweat glands (Fox and Hilton, 1958) entails the quick release of biologically active bradykinin.

Availability of natural inhibitors

In their isolated state, all these endogenous factors, except rabbit protease PF, induce only a transient increase in capillary permeability, lasting at most about fifteen minutes. The plausibility of their role as natural mediators is enhanced if we can find an explanation of this short-lived effect. The transient effect of histamine is explicable in terms of diamine oxidases, acetylating and other enzymes present in the tissues, and to some extent in terms of certain 'histaminopectic' globulins. That of the proteases can be explained by the inhibitory globulins that accompany the profactor in its wanderings from blood to lymph circulation. Among the possible modes of inactivation by a polypeptide factor, further hydrolysis by peptidases is at least plausible, since *in vitro* this process yields oligopeptides with little permeability-increasing potency.

Our examination in the light of the 'plausibility' criteria clearly does not point to any one of the substances—histamine, 5-HT, polypeptide, or protease PF—as the most likely endogenous mediator of increased permeability, except in the negative way that the polypeptides perhaps lack sufficient potency; that 5-HT is virtually impotent in certain species; and that the stimuli needed for the release of histamine are rather more radical than those which activate the serum profactor. It is at this point worth reiterating that there is nothing to link the protease PFs with leukotaxine-like polypeptides in one process of mediation. Not only has PF no frank proteolytic activity, but as a permeability factor it is largely insusceptible to inhibition by antihistamine drugs—a fact difficult to reconcile with the notion that it acts *in vivo* by forming leukotaxine, which is itself a histamine-liberator.

Mimicry

In one respect mimicry of the natural response by a proposed mediator is already established, since we are discussing only substances which increase permeability and (presumably) stickiness of the capillary wall. There is no evidence that any of them promote the 'liquidation' of ground substance. They have all been implicated in the induction of the other outstanding feature of inflammation—tissue leucocytosis; in our hands small doses of histamine are but feebly active, the proteases moderately and leukotaxine certainly so. Whether this activity depends on a true chemotactic effect in addition to facilitating diapedesis through the endothelial wall is doubtful (Harris, 1954).

We must conclude, then, that all the candidates except 5-HT qualify for serious consideration as natural mediators.

THE PROOF OF MEDIATION

The criteria of proof, however, are difficult to apply, because no proof is acceptable without a demonstration of a simultaneous, or near-simultaneous, appearance of newly formed mediator and the vascular response—a demonstration that depends on how accurately we can chart the time-course of the events. We cannot always observe the events on a microscopic scale, and in any macroscopic inflammatory focus we choose to examine we are likely to be dealing with a number of our inflammatory 'units' at different stages in their response. The fact that we have to work with a tissue displaying the average [illegible] being content [illegible] es at *about* the [illegible] with the kind and degree of inflammatory stimulus, though when we confine our examination to the results of minimal effective stimuli, there is a certain pattern of response common to many kinds of inflammation.

The time-course of permeability change in injury

Figure 2 summarizes the time-courses of the permeability effects induced by a number of permeability factors, and by various

kinds of injury. All were determined in guinea-pig skin except curve D, which was determined in rabbit skin. The curves record either the intensity of exudation of dye-stained protein, or the diameter of the area of exudation at the treated site when the dye is given intravenously to an animal bearing lesions of the appropriate age in the skin. The responses are plotted for convenience on three different time-scales, the top scale spanning one hour, the middle two spanning ten hours, and the bottom scale one hundred hours.

We may begin with an example of natural inflammation induced by acute bacterial infection. The response (curve G) is biphasic, the first transient phase occurring immediately after infection, the second rising to a peak in the third and fourth hour. The immediate response is probably accidental because it can be largely eliminated by prolonged washing of the bacterial suspension used for inoculation. The second phase is a constant feature of a number of acute infections (Burke and Miles, 1958) and it coincides roughly with the period of frank immigration of leucocytes into the tissues. Curve E summarizes the response to mild thermal injury (54° for 20 seconds) of the skin. It too is biphasic, but, unlike that in the bacterial infection, the immediate phase is a consistent feature of the response (Wilhelm and Mason, 1958). Curve C represents a chemical burn, made by xylol—an immediate increase in permeability, but one which slowly decreases over several hours. Curve F is the monophasic late response to the α-toxin of *Cl. welchii*, and curve I the similar but more delayed response to the α-toxin of *Cl. oedematiens* (Elder and Miles, 1957). The biphasic response H is induced by a short exposure of the skin to ultraviolet light.

It will be seen that from C to I, which exemplify natural kinds of inflammation, the delayed prolonged response is the rule, with a latent period in all but C. The immediate response occurs only after injury by bacteria, xylol, heat and ultraviolet light. When tests are made in the rat, it is noteworthy that although the second phase is also characteristic of all the responses we have so far measured, the immediate phase is lacking, e.g. after injection of bacteria, and is only inconsistently and feebly elicited by thermal injury.

The first phase, when it occurs, has a time-course like all but one of the endogenous mediators we are considering. Curves A and B are the responses to histamine and protease PF, and would serve equally well for 5-HT and leukotaxine, tested in guinea-pig, rat or rabbit. The exception is the protease PF of the rabbit, which in small doses elicits only a delayed response (curve D) in the rabbit skin. (Curiously enough, in rat and guinea-pig skin the responses to rabbit protease PF, and in rabbit skin the responses to rat and guinea-pig protease PF, are all immediate, as in curve A.) This exception apart, for an explanation of the delayed and prolonged second phase we must either look for mediators with a more durable action, or suppose that if short-acting mediators are responsible, there is not only a prolonged period of continuous release, but a continuing susceptibility of vessels to their action.

Detection of mediators in vivo

Coming now to the detection of mediators in action, we find that all the substances in our list have at one time or another been found in inflamed tissues or in exudates. Thus Spector and Willoughby (1957b) found both histamine and 5-HT in early pleural exudates induced in the rat by turpentine; and Menkin, of course, first discovered leukotaxine in late turpentine exudates—a discovery that may well be relevant to the pathological state of the tissues at this stage of severe inflammation, but is not necessarily relevant to the early hours covered by the first and second phases of inflammation, during which, as far as I am aware, no one has ever found an active polypeptide in substantial quantity. The demonstration of activated protease PF in inflamed tissue is all too easy, because the profactor is there in any case, and is so readily activated by laboratory manipulations. Spector (1956) found pro-PF in 6–12-hour protein-rich exudates induced in the pleural cavity by turpentine—a result to be expected in circumstances where copious late exudates tend towards the same composition as serum itself. On the other hand, his demonstration of *active* PF in half-hour exudates more nearly approaches the conditions under which direct proof of mediation is convincing. Nevertheless, the

proof is not watertight because one cannot in these circumstances distinguish PF responsible for the exudation from PF activated *after* the exudation of pro-PF.

INDIRECT TESTS

When direct proof of the sequence 'mediator-then-permeability increase' is difficult, we must turn to indirect tests—the effect of depleting the body of reservoirs of its supposed mediator, and the effect of inhibitors.

Depletion

Rats, for example, may be depleted of a large part of their histamine and 5-HT reserves by treatment with histamine-liberators like compound 48/80, or of the 5-HT reserves alone, by treatment with reserpine. The depletion is not complete, but sufficient to diminish the permeability response to the injection of substances known to liberate histamine or 5-HT. In turpentine inflammation of the pleural cavity of depleted rats Spector and Willoughby (1958) observed a decrease in exudation, but only in the first half-hour or so. At the fourth hour, the exudation was as great as that in undepleted animals, indicating that the histamine and 5-HT effects are confined to the early part of the response. Depletion cannot help very much with proteases and polypeptides. It would be impossible to deplete the abundant reservoirs of profactor in the serum—even severe and prolonged infection and radiation sickness have no effect on the profactor content of guinea-pig serum (Wilhelm *et al.*, 1957)—and to remove all the potential sources of polypeptide would be to destroy the animal.

Depletion experiments are, however, unsatisfactory, not only in fact, but in principle If responsiveness to injury is necessary to existence, and the mediators we seek are essential to the vascular reaction, a really effective depletion is likely to be very disturbing to the economy of the animal, if not fatal.

Inhibition

The same objections apply to the use of inhibitors, but in practice they are less important because, whereas depletion and subsequent repletion are usually slow and the animal therefore

subject for some time to the effects of depletion, inhibitor concentrations can often be built up for an acute test before long-term effects are manifest. Our greater concern in such tests is that the inhibitor should act with the maximum specificity in doses so low that non-specific effects, including general intoxication, can be ignored.

(*a*) *By histamine antagonists.* Among the antagonists of histamine, some are highly specific. Either mepyramine or triprolidine (Actidil), for example, decreases the early part of the response to turpentine in the rat, and to xylol in the guinea-pig. As in the depletion tests, the later part of the response is unaffected. Systemic triprolidine abolishes the first phase of the guinea-pig's response to burns, and it has no effect on the 4-hour response to staphylococcal infection in the rat peritoneal cavity (Table 5).

A systematic study, in several species of animal, of the action of antihistamines—and, indeed, of other inhibitors—on representative inflammation is badly needed. All the available evidence points to the participation of histamine in the early stage of inflammation. But before we hail this as a triumph of pharmacological analysis, we must reflect on two facts: (1) in many kinds of inflammation there is no detectable immediate response, and (2) when there is an immediate response, and it is abolished by antihistamines, the ultimate course of the inflammation is scarcely affected by its absence; and in this context the virtual failure of the more specific antihistamines in anti-phlogistic therapy of man endorses the evidence of experimental pathology.

On these grounds, we must provisionally, but quite firmly, relegate the participation of histamine to the class of neutral events as far as the main response is concerned, and probably a nuisance as far as the body is concerned. As a clincher to this argument, I would cite a third fact—not elicited by dubious manipulations of the experimental pathologist—namely, that the cornea, being avascular in the normal state, cannot respond immediately to any hypothetical histamine liberated by injury, yet in its own good time displays all the stigmata of orthodox inflammation.

(*b*) *By protease antagonists.* The antihistamine results are also useful in clearly separating the delayed response—which is the most constant feature of inflammation—as something not mediated by histamine. But that, alas, is perhaps the most striking positive fact about this second phase, because as far as inhibitors are concerned, it has yielded none of its secrets about mediators. It is virtually insusceptible to anti-proteases like soya bean [illegible] response, but then it depresses all permeability effects, whether induced by histamine, 5-HT, polypeptide or protease, and is therefore useless as an indication that any one of them is active.

The failure of the anti-proteases does not exclude the protease PFs quite as firmly as the failure of the antihistamines excludes histamine, because there are in the body so many other proteases which might selectively deviate a protease inhibitor. This is I think a valid reason for hedging. Another reason is the fact that in all instances tested the serum protease is present (but whether as PF or pro-PF is not clear) during the delayed response. We might also plead pharmacological inaccessibility of protease PF to inhibitor, but that is unlikely unless we discard the hypothesis of an extrinsic mediator, and postulate intracellular activation of PF. Failure to inhibit a substance *in vivo* is clearly a less firm proof of its irrelevance in an event than specific inhibition is proof of its participation. But having said this, we cannot on the present evidence do anything but exclude the serum proteases as proved mediators of the second phase.

Other Mediators

It was the sequence of events in early response to micro-injury, it will be recalled, which indicated an extracellular mediator. We have no comparable observations of the sequence of later events, so that this indication is not necessarily relevant to the problem of the delayed response, 1–4 hours after injury. At this stage, the increase in vascular permeability may be the expression of an intrinsic change, for example, in the metabolism of the capillary endothelium. Again, bearing in mind the duration

of this delayed phase, an extravascular mediator may be a long-acting substance quite unlike those (Fig. 2, A and B) we have already discussed. The delayed phase, moreover, differs strikingly from the early in another respect; in this phase neutrophil leucocytosis of the tissues has at least begun and is often well established. The tissues have gained a new cell, with its own peculiarities of metabolism.

In mature inflammation, tissue respiration tends to be replaced by glycolysis, with a consequent drop in pH, and accumulation of lactic acid. Leucocytes themselves can be strongly glycolytic, and lactic acid itself is a permeability factor about as potent as some preparations of leukotaxine; fifty micrograms suffice to increase permeability of the capillaries in a gram of guinea-pig skin. Now as few as a thousand million polymorphonuclear cells under optimum conditions of anaerobic glycolysis can in one hour produce about 10 μg. of lactic acid (Barron and Harrop, 1929) Given circumstances, then, of not too much buffering, and not too much removal of the lactic acid formed, a pericapillary mass of leucocytes might well raise enough free acid for a local and sustained increase of permeability. The lactic acid notion is worth investigating, but I doubt whether it will prove to be correct. I mention it as an obvious example of a product of the infiltrating leucocyte which might be relevant, and because there is good experimental evidence of a permeability change dependent on the presence of leucocytes. The severe inflammation in the skin of animals in which either the Arthus or the Schwartzman phenomenon has been elicited has the time-course of our typical delayed response. At this stage both infiltration by neutrophil leucocytes and oedema are pronounced. But as Stetson and Good (1951; Stetson, 1951) in the United States and in this country Humphrey (1955) have clearly shown, when the neutrophils of the blood are considerably decreased by treatment with nitrogen mustard or anti-neutrophil serum, the Arthus and the Schwartzman responses are greatly diminished, and both neutrophil infiltration and oedema are absent; only a late accumulation of lymphocytes occurs. The experiments were so conducted that a direct effect of the depleting agents on the endothelium was excluded, and we can

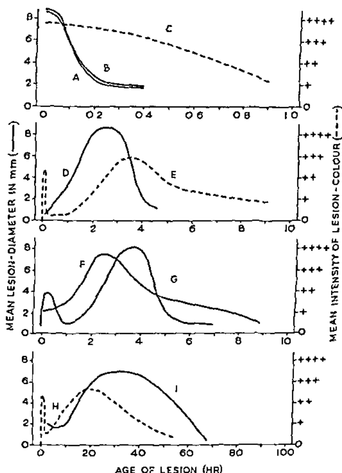


FIG. 2. The time-course of increased permeability of the blood vessels of the skin, after various stimuli. All the curves were elicited in the guinea-pig, except D, which was elicited in the rabbit.

A *Histamine, intracutaneous*
B Protease PF from guinea-pig serum, intracutaneous
C *Xylene, 2 minutes' application*
D Protease PF from rabbit serum, intracutaneous
E Thermal injury, 54° for 20 seconds
F *Cl. welchii* α-toxin, intracutaneous
G Acute bacterial infection, intracutaneous
H Ultraviolet light injury, Kromayer lamp, 20 seconds
I *Cl. oedematiens* α-toxin, intracutaneous

reasonably infer that the oedema is in some way a consequence of some stage in the polymorphonuclear leucocytes' passage from the blood to the tissues.

Conclusion

It will be seen from this analysis that we cannot say of any part of the minimal inflammatory response that it is necessary for ultimate recovery of the injured tissue, and can only guess that most of it is neutral or an unwanted nuisance. We can, however, suggest that the essential part of the inflammatory reaction lies in the delayed responses, and that the immediate response is either neutral or a nuisance. But that essence, embodied in the delayed response, will submit to no generalization about mediators. Nevertheless the analysis, and especially the dissection of various responses into as precise a time-course as possible, indicates the time and place to look for the answers to the question 'Is all inflammation a nuisance, and will knowledge of its mediators help to control it?' The pharmacological dream of an isolable mediator may be unrealizable because the mediator is produced so fleetingly or in such small quantities that it is undetectable in samples from inflamed tissue. And if the dream is realized, it may fade because no inhibitors are specific enough to prove its participation in natural inflammation. There are plenty of other lines of sttack. We might, for example, look for a subtle or perhaps not-so-subtle derangement of tissue respiration, or of a sodium-potassium balance modifying transport through the endothelial cytoplasm. We might concentrate on the general anti-phlogistic effect of, say, salicylates, and its relation to the uncoupling of oxidative phosphorylation; or, recalling that *Cl. welchii* lecithinase is a permeability factor, look into fat metabolism, and its connection with derangements of lipoprotein cell surfaces. Then again, the tissue polysaccharides have been implicated in the induction of tissue leucocytosis (Meier *et al*, 1956), and nucleosides as histamine-liberators that may be released by damaged cells (Spector and Willoughby, 1957a). There is plenty going on in these fields, but as they have not yet illuminated the obscure mysteries of inflammation, I can say little about them.

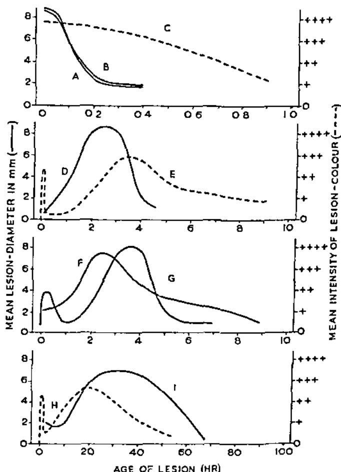


FIG. 2 The time-course of increased permeability of the blood vessels of the skin, after various stimuli All the curves were elicited in the guinea-pig, except D, which was elicited in the rabbit

- A Histamine, intracutaneous
- B Protease PF from guinea-pig serum, intracutaneous
- C Xylene, 2 minutes' application
- D Protease PF from rabbit serum, intracutaneous
- E Thermal injury, 54° for 20 seconds
- F *Cl welchii* α-toxin, intracutaneous
- G Acute bacterial infection, intracutaneous
- H Ultraviolet light injury, Kromayer lamp, 20 seconds
- I *Cl oedematiens* α-toxin, intracutaneous

MENKIN, V. (1956) *Biochemical Mechanisms in Inflammation*. Thomas, Springfield, Illinois.

[illegible]

J. exp. Path. 39, 343.

PAPPENHEIMER, J. R. (1953). *Physiol. Rev* 33, 387.

PASKHINA, T. S. (1956). *Clin. chim. Acta*, 1, 24.

PEASE, D. C. (1955). *Anat. Rec.* 121, 701.

RAWSON, R. A. (1943). *Amer. J. Physiol.* 138, 708.

RHODIN, J. (1955). *Exp. Cell Res.* 8, 572.

RILEY, J. F. and WEST, G. B. (1953). *J. Physiol* 120, 528.

ROWLEY, D A. and BENDITT, E P. (1956). *J. exp Med.* 103, 399

SHIRLEY, H. H., WOLFRAM, C. G., WASSERMAN, K. and MAYERSON, H. S (1957). *Amer. J. Physiol.* 190, 189.

SMITH, D. D. and MILES, A. A. (1960). *Brit. J. exp. Path* 41, 305.

SPARROW, E. M. and WILHELM, D L. (1957). *J. Physiol.* 137, 51.

SPECTOR, W. G. (1951). *J. Path. Bact.* 63, 93

SPECTOR, W G. (1956). *J. Path. Bact.* 72, 367.

SPECTOR, W G [illegible]

SPECTOR, [illegible]

SPECTOR, [illegible]

SPECTOR, [illegible]

STETSON, C A. (1951). *J exp. Med* 93, 409, 94, 347.

STETSON, C. A. and GOOD, R. A. (1951). *J. exp. Med.* 93, 49

UNGAR, G (1956) *Histamine. Ciba Foundation Symposium.* Churchill, London, p 341

WILHELM, D. L. and MASON, BRENDA (1958) *Brit. med. J.* ii, 1141

WILHELM, D L, MILES, A. A and MACKAY, M. E. (1955). *Brit. J exp Path* 36, 82.

WILHELM, D L, MILL, P. J and MILES, A. A. (1957) *Brit J. exp. Path* 38, 446.

WILHELM, D L, MILL, P. J, SPARROW, E. M, MACKAY, M. E and MILES, A. A (1958) *Brit J exp Path.* 39, 228.

ZWEIFACH, B W (1953). In *The Mechanism of Inflammation* Acta Inc., Montreal, p 77

ZWEIFACH, B. W. and CHAMBERS, R. (1950). *Ann. N.Y. Acad. Sci.* 52, 1047.

As I intimated at the outset, there are no clinical applications. Indeed, the burden of my lecture may have seemed reminiscent of Albert Chevalier's old song 'Wot's the good of Hanyfink? Why—Nuffink.' But I contend it is not as nihilist as all that. The imposing array of negative results, if accepted, has, I hope, cleared the field of some of its speculative litter, and at least holds out a promise of further advance. I hope too that I have conveyed some of the excitement of chasing substances that look promising as mediators. I have probably imbued all who read this lecture with the sense of the disappointment, as each cherished notion comes to rigorous test, in learning once again that a substance can have an effect without having a discoverable function; and I have certainly imposed on them the exhaustion that seizes those who try to cover this vast and intractable field.

REFERENCES

Armstrong, D., Jepson, J. B., Keele, C. A. and Stewart, J. W. (1957). *J. Physiol.* 135, 349.

Barrow, E. S. G. and Harrop, G. A. (1929). *J. biol. Chem* 184, 89.

Becker, E. L., Wilhelm, D. L. and Miles, A. A. (1959). *Nature, Lond.* 183, 1264

[illegible] *Bact.* 76, 1.

[illegible] *J. Anat.* 57, 385.

[illegible] *J. exp Path* 20, 417.

[illegible] 133.

[illegible] *th.* 39, 335.

[illegible]

[illegible] 19.

Friedberger, E. (1909). *Z. ImmunForsch* 2, 208.

Gorkin, V. Z. (1957). *Clin. chim. Acta*, 2, 83.

[illegible]

[illegible] *Physiol.* 134, 471.

[illegible] 36, 268, 283

[illegible] *J. exp. Med*, 19, 480

[illegible] 98, 65

[illegible]

[illegible]

[illegible], 1041

Meier, R., Desaulles, P and Schär, B. (1956). *Verh. Naturf. Ges. Basel*, 67, 447.

The normal pressure in the pulmonary circulation is about 16/7 mm. Hg. (Wood, 1958) and hypertension is said to exist when the pressure is about double this—30/15 mm. Hg. In the systemic circulation there is probably only one mechanism for the production of hypertension, namely increased arterial or arteriolar tonus but this is not true of the pulmonary circulation. Here hypertension can be produced by a variety of mechanisms including simple mechanical blockage. The mechanisms of production of pulmonary hypertension may be classified as follows:

- Passive
- Hyperkinetic
- Organic vascular obstruction
- Increased vascular tonus
- Mixtures of the above.

Passive pulmonary hypertension

In this form there is a back pressure from the left atrium impeding the outflow from the lungs. Examples are:

- Left ventricular failure
- Mild mitral stenosis or mitral incompetence
- Left atrial myxoma
- Pulmonary venous or atrial thrombosis.

To keep up an adequate pulmonary blood flow there needs to be a pressure gradient of about 10 mm. of mercury across the lungs and therefore to produce a pulmonary arterial pressure of 30 mm. Hg. the left atrial pressure would need to be about 20 mm. Hg. Theoretically this pressure ought not to be exceeded or the pulmonary capillary pressure would rise above the colloid osmotic pressure and pulmonary oedema would result. In practice higher left atrial [illegible]
pulmonary [illegible]
with consid [illegible]
ever, unusual and in general passive pulmonary hypertension is of minor degree.

XIII

Pulmonary Hypertension

C. V. HARRISON

THERE are differences in structure and function between the systemic and pulmonary vasculature and these have a bearing on the production of hypertension in the two circulations. The pulmonary vasculature has relatively little need to alter the distribution of blood within the lungs by directing more or less to different parts according to the needs of the moment. Furthermore, it does not have to counteract the effects of gravity to any degree. On the other hand it has to transmit the total cardiac output through a limited anatomical territory. The pulmonary vasculature must therefore be capacious but can and does have a low pressure, and this is reflected in its structure. Elastic arteries, which have little contractility, extend from the right ventricle down as far as the end of the cartilaginous bronchi—to a diameter of about 1 mm.: this is in striking contrast to the systemic circulation where elastic arteries end at the level of the carotids or the subclavians. In the lungs muscular arteries have a short range roughly corresponding with the bronchioles and extending from a diameter of about 1 mm. to about 0·1 mm. The size range in the arterioles is not much different from the systemic circulation. At all levels the pulmonary arteries are much thinner than the corresponding systemic ones and this difference is due to a smaller complement of muscle; the fibrous and elastic components being little altered. This applies only to post-natal life: during foetal life the pulmonary arteries are as thick-walled as the corresponding systemic arteries.

is due to vascular contraction resulting from anoxaemia but loss of vascular territory plays some part.

Tonic vascular contraction

Primary pulmonary hypertension
Secondary to passive pulmonary hypertension
Secondary to hyperkinetic pulmonary hypertension
Secondary to anoxaemia
Pulmonary collapse.

These are the conditions that cause a rise of pressure in the pulmonary circulation: what are the effects that they produce in the pulmonary vasculature?

For some years I have studied the lungs from different types of pulmonary hypertension by postmortem angiographs followed by histological studies and it is the results of these investigations that I wish to put before you.

[illegible]

therefore discuss in this group all the cases I have examined in which the pulmonary hypertension was due to back pressure from the left atrium. This list includes 40 cases of mitral stenosis and 12 other examples of passive pulmonary hypertension.

In discussing the vascular pathology in this group it will be easiest to describe first the changes in a typical case of mitral stenosis and then to deal with the variations from this.

The postmortem angiograph of a typical mitral stenosis is characteristic (Plate VI, Fig. 1):[1] the pulmonary artery at its entry into the lung is dilated in approximately half the cases and normal in the other half. The segmental arteries vary between the upper and lower zones of the lung. In the upper zones they are usually normal-sized but sometimes dilated; in the lower zones they are either normal-sized or narrowed about equally often. There is a perceptible difference between the two zones in about two-thirds of the cases.

[1] The plates referred to in this lecture will be found between pages 232–3.

Hyperkinetic pulmonary hypertension

This group includes all the cases in which the pulmonary artery receives blood in either an unduly great volume or at an unduly high pressure. It includes therefore atrial and ventricular septal defects and communications between aorta and pulmonary artery. In atrial septal defects there may be a high flow and low resistance without necessarily a rise of pulmonary pressure although the latter usually occurs in time. In ventricular and aorto-pulmonary communications there is inevitably a transmission of left ventricular pressure into the pulmonary artery. In the latter cases the pulmonary vasculature must offer a high resistance otherwise the peripheral circulation would not receive enough blood for the body's needs.

Organic vascular obstruction

The lesions included under this heading can be classified as follows:

Within the vascular lumen:

- Thrombi
- Tumour emboli
- Parasites
- Fat and air emboli (temporarily)

in the vessel walls:

- Endarteritis
- Polyarteritis
- Scleroderma

outside the vessels:

- Emphysema
- Pulmonary fibrosis
- Pneumonectomy.

In addition to these pulmonary collapse causes a local closing down of the vasculature and therefore in effect a loss of territory though the mechanism is that of muscular contraction. It is generally accepted that in emphysema most of the hypertension

thrombotic origin. In favour of this is an association between thrombo-embolism and occlusive atheroma and even more convincing was the finding of mural thrombi being incorporated into atheromatous patches.

In cases of mitral stenosis thrombo-embolism is a frequent and serious complication. It was present in 20 out of 40 cases and in the lungs actually examined by angiography there was an average of over four (4·3) occlusions per lung in these cases. The site of the occlusions was interesting. Out of 126 occlusions there were 24 in the upper lobe, 30 in the middle lobe and 72 in the lower lobe. This distribution corresponds with the distribution of emboli. These emboli play a significant part in the patients' deterioration. They cut off territory in an already burdened circulation and due to infarction they produce pleural effusions and further pulmonary collapse.

Passive venous congestion

Venous congestion appears to reflect the mode of dying. In patients dying in congestive failure or cardiac cachexia the degree of congestion was usually not particularly severe. The typical severe congestion was seen almost entirely in patients dying in acute pulmonary oedema. In the latter it was noted that oedema and congestion was maximal in the upper and middle zones and that the lower zones often escaped (Plate VIII, Fig. 3). It is possible that the more hypertrophied arteries in the lower zones helped to protect these parts from capillary engorgement

Kerley's 'B' lines

It is now generally agreed that the horizontal linear shadows described by Kerley (1933 and 1951) are due to oedematous thickening of the interlobular septa where they happen to lie in the plane of the X-ray. We have been able to confirm the presence of such oedematous septa in cases of passive pulmonary hypertension (Plate IX, Fig. 4).

These are the findings in typical cases of mitral stenosis. In other forms of passive pulmonary hypertension the lesions were similar but usually of considerably less severe form.

Coming to the smaller pulmonary arteries this tendency was exaggerated, and there was a difference between upper and lower zones in about three-quarters of the cases.

On microscopical examination it was found that this lower zone narrowing was accompanied by and presumably due to exaggerated arterial hypertrophy (Plate VII, Fig. 2). It is interesting that such hypertrophied arteries retained their tonus after death sufficiently to produce these characteristic angiographs. This could be shown by perfusing such lungs with a solution of sodium fluoride which caused loss of muscular contraction and loss of the characteristic pattern.

If as we have suggested the pulmonary hypertension is due to a raised left atrial pressure there ought to be a corresponding hypertrophy of pulmonary veins and even perhaps differential hypertrophy. In fact this is found. It seems reasonable to suggest that the difference between zones in this may be due to gravity since the absolute pressures are so much lower. If so, perhaps the greater contraction and hypertrophy on the arterial side may be dependent on some reflex connection between arteries and veins.

Vascular complications of mitral stenosis

It has been known since the work of Moschcowitz in 1927 that pulmonary hypertension caused exaggerated pulmonary atheroma and it is natural to wonder whether the narrowing seen in these angiographs or in corresponding *in vivo* ones could be due to atheroma encroaching of the lumen. Briefly one can say that in the cases we have so far described atheroma is not the cause. It is, however, the cause of a different picture, namely that of beaded or localized narrowing. This was seen in about one-third of cases and was always confined to the lower zones. One should perhaps mention here that pulmonary atheroma is ordinarily a very mild lesion even in cases of pulmonary hypertension. To find therefore in some cases of mitral stenosis atheroma as severe as that seen in coronary arteries and capable of causing gross narrowing is a matter for surprise. This in turn leads one to wonder whether this gross type of pulmonary atheroma may not be as Duguid (1946) has suggested, partly of



Fig. 1 Post-mortem angiograph from a case of mitral stenosis The arteries in the upper lobe and in the apical segment of the lower lobe appear normal, those in the middle lobe and in the basal segments of the lower lobe are narrowed

Hyperkinetic pulmonary hypertension

In this group I have been able to study 20 cases. They can be divided into two main groups: those with a high flow typified by the atrial septal defects and those with a high pulmonary resistance and low flow typified by the ventricular septal defects.

Atrial septal defects

The typical angiographic finding here is that of enormous dilatation of the whole arterial tree down to the end of the elastic arteries (Plate X, Fig. 5). If there is a high flow and low resistance and therefore low pressure then the muscular arteries are of normal size. If, as is usual in cases coming to autopsy, there is high resistance and high pulmonary arterial pressure then the muscular arteries undergo first hypertrophy and then intimal fibrosis and this produces in the angiograph a loss of filling in small arteries. In contrast to what we have seen in mitral stenosis there is no zonal difference that we have been able to detect, either in the angiographs or in the histology.

Ventricular septal defects

In these cases the angiographs show only a mild degree of dilatation of the elastic arteries but there is a striking, sudden loss of vascular filling at the muscular artery level. Microscopically there is generalized muscular hypertrophy which is more severe than in atrial septal defect, together with a severe degree of intimal fibrosis.

Organic vascular obstruction

Intravascular. Pulmonary thrombosis and embolism in its common form does not cause persistent pulmonary hypertension but in a minority of cases the emboli are small and obstruct sufficient of the circulation to cause hypertension but not enough to [illegible]

etiology it is essential to make a detailed dissection of the pulmonary arteries and also to examine many sections.

Malignant tumours can also occlude the pulmonary vasculature sufficiently to cause hypertension.

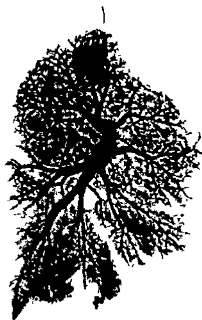


FIG. 1. Post-mortem angiograph from a case of mitral stenosis. The arteries in the upper lobe and in the apical segment of the lower lobe appear normal, those in the middle lobe and in the basal segments of the lower lobe are narrowed

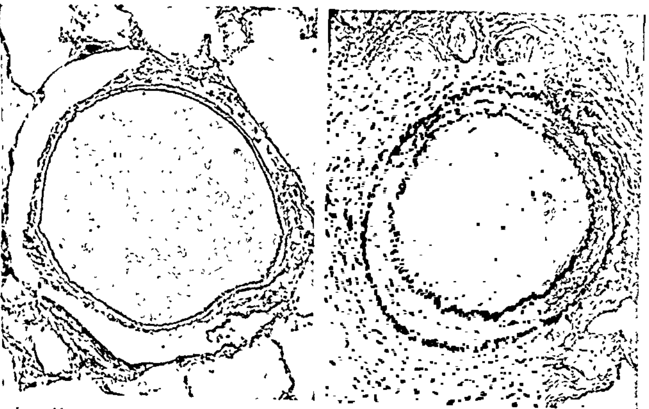


Fig. 2 Mitral stenosis (same case as Fig. 1) Upper (*left*) and lower (*right*) lobe muscular arteries cut at levels 1 cm. from apical and basal pleural surfaces The upper lobe artery is hypertrophied the lower lobe artery is much more hypertrophied (Both × 110)

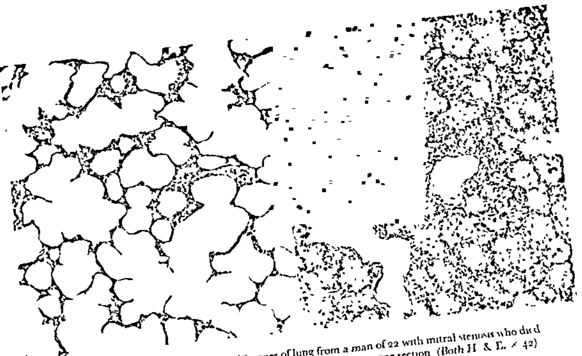


FIG. 3 Upper (*left*) and lower (*right*) zones of lung from a man of 22 with mitral stenosis who died of acute pulmonary oedema. Congestion is seen only in the upper zone section. (Both H. & E. × 42)

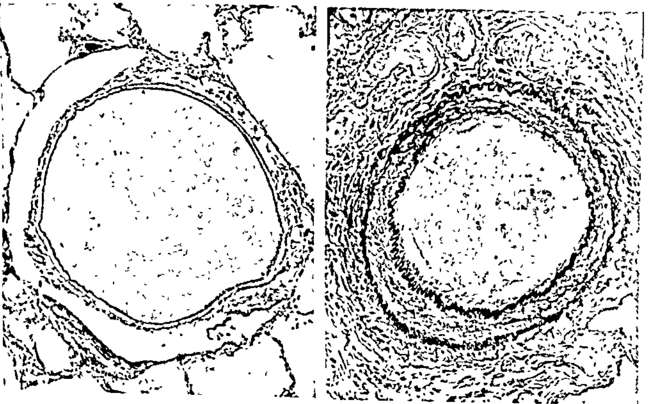


FIG. 2. Mitral stenosis (same case as Fig. 1). Upper (*left*) and lower (*right*) lobe muscular arteries cut at levels 1 cm. from apical and basal pleural surfaces. The upper lobe artery is hypertrophied, the lower lobe artery is much more hypertrophied. (Both × 110)

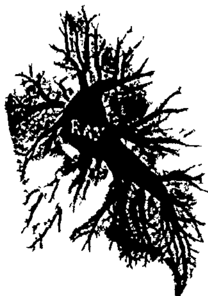




FIG 5 (*left*) Post-mortem angiograph from a woman of 45 with an atrial septal defect The large arteries are greatly dilated, the small peripheral branches are narrowed

FIG 6 (*right*) Post-mortem angiograph from a case of primary pulmonary hypertension The main pulmonary artery and its immediate branches are of normal size but distal to this there is a great diminution in size of the arteries maximal in the small branches

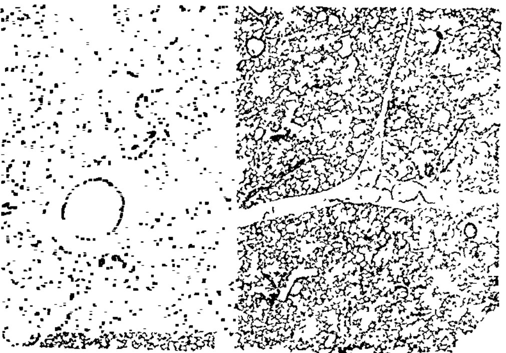


FIG 4 Mitral stenosis Section of apical segment showing widened interlobular septa containing injected veins and distended lymphatics (H. & E. × 6 5)

Permeation of the periarterial connective tissue by cancer cells leads to strangling of the arteries in cases of so-called lymphangitis carcinomatosa. In addition embolism of the arterial lumen by tumour can cause fatal pulmonary hypertension. I have seen one case due to chorion epithelioma and Dr. A. Rickards of Lancaster has shown me another due to carcinoma of the stomach.

Parasitic embolism occurs with Bilhargia infestation in Egypt and Puerto Rico. Infestation is extremely common, pulmonary embolism by parasitic ova occurs but is slight and only very rarely involves sufficient arteries to cause hypertension (Shaw and Ghareeb, 1938).

Disease of the arterial wall

We have already mentioned endarteritis or intimal sclerosis as a reaction to hypertension. Polyarteritis nodosa only involves the pulmonary circulation in a minority of cases and even then is not a significant cause of hypertension. One arterial lesion that may cause pulmonary hypertension is scleroderma. I have examined one case Here the angiography did not show any detectible change but microscopically there were lesions in the smallest muscular arteries and the arterioles which I presume were significant since the patient died of right heart failure and there was hypertrophy of the larger muscular arteries.

Finally, of the obstructions due to lesions outside the vessels, the only one we have had the opportunity of studying is emphysema. In this disease the elastic pulmonary arteries are dilated, distal to this the picture is variable depending on the distribution of the emphysema. Where there is centrilobular emphysema there is loss of territory at the level of the small muscular artery or arterioles. On the whole the loss of territory is very slight and probably does not account for much of the pulmonary hypertension seen in this disease.

Tonic vascular contraction

The forms of pulmonary hypertension in which tonic vascular constriction is secondary to some other change, such as mitral stenosis, have already been considered. The only form in which

XIV

Chronic Bronchitis and Hypersecretion of Mucus

LYNNE REID

UNTIL recently chronic bronchitis might have been thought a somewhat commonplace subject to be included in a series of lectures on the scientific basis of medicine. Even today it feels a little self-conscious, for it has long been a Cinderella of diseases, but because of the suffering and disability it causes, not to mention the loss of working time, its social and economic significance for the community has brought it into the limelight. Moreover, to the extent to which acute infectious diseases are conquered and brought under control, so perforce chronic bronchitis is seen in higher relief.

Behind the diagnosis 'chronic bronchitis' are unanswered questions of the normal and abnormal functioning of the bronchial tree, particularly in relation to secretion of mucus—wherein lies the secret to the understanding of the condition. The crux of this problem is to ascertain the causes of excessive secretion, a persistent feature of the disease from its onset to the stage of grave disability. Recurrent infection, which plays so large a part in producing that disability, presents a further aspect—the interaction of hypersecretion of mucus with infection. Discussion of chronic bronchitis reasonably begins with a description of the structures responsible for secretion and of the factors which are known to influence its production.

Structures Responsible for Secretion of Mucus

Mucus is secreted by cells in the surface epithelium lining the bronchial tree and by branched tubulo-acinar glands lying in

tonic muscular contraction is believed to be the sole cause is in so-called primary pulmonary hypertension. Some workers hold that such a condition does not exist and that all cases have some other explanation. Clinically it is virtually impossible to make a certain diagnosis because of the impossibility of excluding embolism or other rare causes. Pathologically it is nearly as difficult. I think angiography is essential in order to examine the whole pulmonary vasculature properly. With this and with extensive histology one can exclude virtually all the known causes.

I have been able to study two cases in this way. These were women between 20 and 30 years with histories of dyspnoea and progressive right ventricular failure of one to two years' duration. In both patients the postmortem angiographs showed characteristic pictures (Plate X, Fig. 6). The vascular outlines were approximately normal down to the distal part of the elastic arteries; beyond this there was a sudden striking diminution of vascular size with almost complete failure of penetration beyond the proximal parts of the muscular pulmonary arteries.

Microscopically, both cases showed striking hypertrophy down to the levels of the smaller muscular arteries and beyond this the type of change seen in ventricular septal defect. In these two cases I think it is highly probable that the mechanism of hypertension was one of spastic vascular contraction.

ACKNOWLEDGEMENT

We are indebted to the Editors of the *British Journal of Radiology* for permission to reproduce all six illustrations.

REFERENCES

DUGUID, J. B. (1946). *J Path Bact.* 58, 207.
KERLEY, P. J. (1933) *Brit. med J.* 2, 594.
KERLEY, P. J. (1951) *A Textbook of X-ray Diagnosis.* H. K. Lewis, London, II, 404.
MOSCHCOWITZ, E. (1929) *Am J med Sci.* 178, 244
BERNARD SHAW, A. F and ABOU GHAREEB, A (1938). *J Path. Bact.* 46, 401.
WOOD, P. (1958). *Brit. Ht J.* 20, 557.

described a cell to which he gives the name 'cellule en brosse' and which, he suggests, may be of importance in humidification.

The distribution of mucous glands roughly parallels the distribution of plates of cartilage in the wall of the bronchi, but not completely, as cartilage may extend further distally along the bronchial tree than do the glands. Because of its small size, it is easier to examine the foetal lung in serial and step sections and in my department we have, therefore, used embryological material to ascertain the distribution and numbers of glands (Bucher and Reid, in press). The interpretation of certain appearances in the adult have been made easier thereby. The somewhat laborious task of estimating the number of glands is [illegible]

Cartilage and mucus glands appear first in the trachea and then progressively further out in the bronchial tree—the cartilage appearing a little ahead of the [illegible]

segmental bronchi. The period of greater growth is from the twentieth to the twenty-fifth week; thereafter the increase is slight. [illegible]
no mor [illegible]
an axi [illegible]
in the epithelium, from which tubules, simple or branched, grow in the wall. They are recognizable as glands before any material giving positive stains for mucus is found in the cells. Such material soon appears, however, though rarely in surface epithelium in foetal life. The glands pass deeply into the wall and beyond the cartilage, so that soon after they first appear they are found external to cartilage. Thus such an appearance in the adult is not in itself evidence of hypertrophy.

To give some idea of the number of glands along one bronchial pathway they were counted as ducts in serial sections. Taking the superior lingular pathway as an example, at birth the duct concentration was 21 per sq. mm. in the superior lingula bronchus. Absolute figures for successive generations were 71 ducts in the segmental bronchus, 90 in the first division

the depth of the bronchial wall. In both these sites and even in the ducts of the glands are cells which do not stain for mucus and, broadly speaking, two types of cell which do—cells distended with secretion and others containing discrete granules (Plate XI, Fig. 1).[1] The surface cells distended with mucus are generally called goblet cells. In the glands the cells are usually described as either serous or mucous, but this division is misleading for two reasons—first because it is based on the haematoxylin-eosin appearance of the parotid and submaxillary glands, a classification which has been found unsatisfactory even in relation to the gut, as Babkin (1950) has shown. He is among the authorities who have emphasized the limitation of this oversimple classification, a limitation which applies even more to the bronchial tree. Secondly, with haematoxylin-eosin it is possible to recognize three types of cell in bronchial glands, the first with abundant palely staining basophilic cytoplasm, i.e. mucous; the second and third, usually designated serous, are smaller cells with centrally placed nuclei and cytoplasm, either brightly eosinophilic, or somewhat basophilic. The eosinophilic cells, less common than the basophilic, are seen in both normal and diseased conditions, including chronic bronchitis, and in both the mucosa of the ducts and the acini. Special stains show that mucous cells are distended with mucus, others contain granules, while others again are free; either type of serous cell, the brightly eosinophilic or basophilic, may contain mucus granules. The fundamental significance of these differences is not known. Some authorities, such as Gammarrota (1954) and Leonardelli and his colleagues (1957) believe that the types are distinct, others that they are only different phases of secretory activity (among them Florey, Carleton and Wells, 1932; Policard and Galey, 1945). This latter view seems most likely to be right as, in disease, cells which do not normally secrete mucus can develop the faculty, which emphasizes their capacity for change from one type to another.

In animal tissue electron-microscopy has recently distinguished cell types not previously recognized. Rhodin (1957), for instance, working on the bronchial epithelium of the rat, has

[1] The plates referred to in this lecture will be found between pages 218-9

CONTROL OF BRONCHIAL SECRETION

Our knowledge of the function of mucus-secreting structures, especially the bronchial tree, is limited, depending as it does on deduction from animal experiment. Further, because it is technically more difficult, less work has been done on trachea than on gut. From the classic experiments of Florey, Carleton and Wells (1932) on the trachea of the cat and dog it would seem that the goblet cells of the surface epithelium secrete in response to direct irritation, but that the stimulation of the vagus nerve has no effect on them. The glands, however, seem to be under nervous control since vagal stimulation causes glandular secretion, as does injection of pilocarpine. Moreover, nervous pathways exist between the epithelium and the glands, for stimulation of the surface causes reflex glandular secretion, which can be prevented by atropine (Fig. 5).

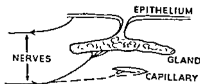


FIG 5 Bronchial wall illustrating nervous control of glands—efferent pathways in the vagus nerve, afferent pathways coming from the surface epithelium

Florey's attempt to separate the secretion of the surface cells from that of the glands in the wall was frustrated by the impossibility of keeping the surface epithelium dry, since it possesses the striking ability to produce fluid without the intervention of the glands. The secretion of the glands can be completely controlled by atropine but when dry air is passed over the tracheal epithelium the same amount of fluid collects in spite of atropinization. From histological studies of the surface epithelium it appears that even the goblet cells play little part in producing the fluid, which probably results in a rapid transudation of fluid through surface epithelium, a possibly important factor in humidification.

Cells do not secrete uniformly and, broadly, the secretion of

beyond, 32 in the next, 11 in the next, gradually decreasing until in the eighth division there were only one or two, a total of 225 in this single axial pathway. Although the total number varies somewhat from foetus to foetus, these figures give an idea of gland distribution. In the main bronchi of the adult the concentration of glands, according to von Hayek (1953), is something like 1 per sq. mm., which figure is consistent with our findings.

In the normal adult lung, goblet cells are about 1 in 4 of the epithelial cells of main bronchi but, like the glands, they decrease progressively towards the periphery, so that in any one cross-section of a normal bronchiole (Plate XI, Fig. 2), for example, no mucus-secreting cell is seen. This means that in finer airways, near the respiratory part of the lung, mucus is not normally produced. At birth goblet cells are still sparse, much fewer, relatively, than in the adult. As the area of the bronchial lining increases so greatly between childhood and adult life the normal quota of goblet cells in the latter may reflect environmental factors in childhood.

RELATIVE IMPORTANCE OF SURFACE AND GLANDULAR EPITHELIUM

What is the relative importance of glands and goblet cells in the adult? A rough idea can be gained by comparing the volume of glandular tissue with that of the goblet cells of the surface epithelium. The former was calculated by multiplying the average thickness of the gland layer by the surface area of trachea and bronchi as far as and including the fifth generation of intra-segmental bronchi. The volumes of the goblet cells is about 0·1 ml., the volume of glands 4 ml.; thus the gland volume is about 40 times that of the goblet cells. Although the volume of secretion cannot be precisely related to that of secreting tissue, broadly speaking it would seem justified to deduce that even in the normal the glands are responsible for much greater mucus secretion than the epithelium. Absolute measurements of the depth of the gland layer are dealt with later in relating the volume of sputum to changes in the glands.

every 24 hours. These results suggest that in man 100–150 ml. in 24 hours is possible (Policard and Galy, 1945). Perry and Boyd investigated the effect on secretion of changes in temperature and found that a lowering of temperature caused a fall in secretion. Our ignorance of the effect on secretion of the contributions of epithelium, glands and transudation respectively, makes us conscious of the dearth of our knowledge. We cannot yet gauge secretion experimentally, let alone in man—which may explain why the experimental-clinical approach of applying drugs to man and measuring the effect on amount and viscosity of sputum have so far proved disappointing (Forbes and Wise, 1957).

Relation of Cilia to Secretion

Any consideration of the secretion of mucus raises the question of its removal. This is effected by what Proetz has called the 'ciliary escalator' (quoted Hilding, 1943); once the secretion is outside the cell and resting on the bronchial epithelium the cilia come into play.

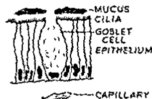


Fig. 6 Diagram showing two layers of secretion on the surface of the epithelium—thick mucus lying on the tips of the cilia, a less viscous layer surrounding them. Epithelial cells free of mucus, containing granules staining for mucus, distended by secretion, or discharging it, the amount of fluid passing through the cell is influenced by the state of the capillaries and hydration of the wall

Lucas and Douglas (1934), studying the nose, have shown, and Hilding (1943) has confirmed, that the cilia are not bathed in mucus, but that outside the cells secretion is identifiable as two layers—a superficial layer of thick mucus resting on the tips of the cilia and beneath this and bathing the cilia a thinner, more fluid, layer (Fig. 6). This latter is the medium in which the cilia beat, and in which their movement causes the strands

each cell can be considered of two types (Babkin, 1950). The first consists of specific substances, actively manufactured, and probably recurs with a natural rhythm relatively independent of external factors, but may also result from surface irritation. The second is more passive—the passing of tissue fluids through the cells; while these may contain certain salts and organic material, the amount is not significant. Hydration of the wall and the degree of dilatation of the capillaries in the wall are, therefore, further factors which may influence the level of secretion (Figs. 5 and 6).

Where there is cardiac failure accompanied by bronchitis the primary factor in hypersecretion is probably this increase in the tissue fluids, a hypothesis supported by observation of patients with mitral stenosis. In this connection a survey by Wood (1954) indicated that 25 per cent of both the surgical and medical groups had attacks of winter bronchitis. Chronic bronchitis is predominantly a disease affecting males (Stuart-Harris and Hanley, 1957; Oswald, 1958) but in cases of mitral stenosis it is common among women. To quote Wood, 'One of the remarkable results of technically successful mitral valvotomy is the disappearance of winter bronchitis', which supports the view that in the majority of cases the critical factor in the excessive secretion which is the basis of the diagnosis of bronchitis is the alteration of the fluid balance in the tissue of the bronchial wall.

Experimentally, secretion from a mucus-secreting surface can be increased by a variety of surface irritants—particulate matter organic or inorganic, bacteria alive or dead, noxious gases and noxious fluids (e.g. Florey, 1930; Florey, Carleton and Wells, 1932).

Amount of Secretion

Briefly, the constituents of secretion have been shown by Binet and Bourlière (1944) to be 95 per cent water and 5 per cent dry residue, of which about half is mucus. The extent to which man secretes mucus is not known, so that it is necessary to fall back on animal experiments. Perry and Boyd (1941) have shown that, in the cat and the rabbit, the bronchial tree secretes continuously even under normal conditions and that secretion of the order of 2 ml. per kilogramme of body weight occurs

known *nor*, when there is hypersecretion, what proportion is coughed up. The hypersecretion giving rise to sputum is reflected in an increase in the mucus-secreting elements, both glands and goblet cells (Reid, L., 1954 and 1958). Plate XI, Figure 3 shows part of a normal main bronchus from a male adult, adjacent and in contrast to part of the main bronchus from a male adult patient with chronic bronchitis. In the latter the total thickness of the gland is increased, the individual acini are bigger and more cells are distended with secretion.

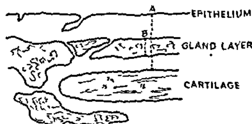


Fig. [illegible] A represents the 'internal [illegible] vary

One of the difficulties in epidemiological studies of chronic bronchitis is the establishing of criteria for diagnosis. Certain diseases can be diagnosed through a specific lesion—a tubercle or an Aschoff nodule—but in chronic bronchitis there is no typical lesion. The changes are only a hypertrophy and, therefore, can only be assessed by comparison with the normal.

In order to reach a quantitative assessment of the degree of change in chronic bronchitis the thickness of the gland layer in bronchi was measured (Reid, L., 1960). The main, lobar, or segmental bronchi were used because the thickness of their walls is similar, although as the diameter of the bronchus decreases the thickness of the wall relative to the lumen is greater. Figure 7 shows the three measurements which were used—A, the thickness of the wall from cartilage to mucosa (i.e., internal thickness) and B, the thickness of gland at the same point

or carpets of thick mucus which lie on the surface to move upwards. Lucas and Douglas have further shown experimentally that the thicker and denser the strands of mucus the more efficient the upward movement.

The efficient functioning of this very delicate mechanism is dependent on *the maintenance of a proper balance between the* two layers, which can be disturbed in several ways. Excessive drying of the mucus may destroy the physical state at the interface between them. Too deep a fluid layer could prevent the mucus secretion on the surface from being in contact with the tips of the cilia; too shallow a serous layer might mean that the mucus would be entangled in the cilia, preventing them from functioning normally—this may occur, for instance, in certain cases of status asthmaticus, where secretion does not become free of the cell (Houston, Navasquez and Trounce, 1953).

Normal Function Summarized

Arguing, then, from examination of the normal structures in the human lung and experimental physiological studies, it would seem that there is a minimum daily secretion from the bronchial tree of perhaps 100–150 ml. This comes from goblet cells in the epithelium and glands in the wall, of which the latter, to judge by volume, are the more important. Glands are under nervous control, goblet cells are not, but the presence of afferent pathways from the epithelium means that the glands are affected by surface irritation. In addition to the mucoid material from the glands and goblet cells there may be transudation of fluid *through the cells of the epithelium, in which the state of the* capillary circulation and the degree of hydration of the wall are important. The functioning of the cilia is dependent on the maintenance of a proper balance between the constituent layers *of the secretion, particularly as regards their amount and* viscosity.

Changes in the Bronchial Wall in Chronic Bronchitis

Glands. The characteristic feature of chronic bronchitis is the hypersecretion which manifests itself as sputum. What increase beyond the normal is necessary before sputum is produced is not

increase was seen in all cases of chronic bronchitis, it suggests that hypertrophy of glands occurs in all the large bronchi.

From thirty counts on ten lobar bronchi of seven operation cases of bronchiectasis a similar increase is seen, the mean for the group being significantly different from the normal. The average ratio of gland thickness to wall thickness was 0·54. If the success of local resection in stopping sputum production is evidence, gland hypertrophy in this condition may be local.

The figures from which these averages were compiled show little variation *inter se*; the normal ratio varied from 0·14 to 0·36. In only four of the bronchi from the cases of bronchitis was the ratio less than 0 5, the maximum being 0·79. In the bronchiectasis group the range of ratios was wider, from 0·33 (the normal) to 0 80. This fits with the clinical history of these patients, as great deformity of peripheral bronchi is not necessarily associated with large amounts of sputum. In some patients there was often only a trace of sputum, while in others several ounces were still being produced although the patient was taking antibiotic drugs continuously.

The figure in the last column in Table 1 represents the number of acini seen in one high-power field. Fewer acini, an average of 14 instead of 24, are seen in a chronic bronchitic than in the normal, and similarly with bronchiectasis, 16 as against 24. This difference results from an increase in the diameter of the acini in chronic bronchitis, which also encroach on the inter-acinar connective tissue. The increase in the depth of the glands is relatively greater than the increase in the diameter of an acinus, suggesting that there is hyperplasia as well as hypertrophy—new acini, not merely an increase in the size of the pre-existing ones. The cells particularly increased are those distended with mucus. These form the majority, but there are still occasional cells in which granules can be seen and some which are free of them.

These results show a clear distinction between normal subjects and the group of cases of established bronchitis. To investigate whether with lesser amounts of sputum there is a correlation between the amount produced and the degree of gland hypertrophy, similar measurements were made in 37 bronchi

(they were measured in units of a graticule placed in a ×6 eyepiece with a 16 mm. objective). These are expressed as a gland/wall ratio. The third measurement was made with a 4 mm. objective and represents the total number of acini which could be seen in one field of 0·4 mm. diameter.

TABLE 1. Measurements of bronchial wall and glands in bronchi from normal subjects and from patients with chronic bronchitis, bronchiectasis, and emphysema without sputum

	No. of cases	No. of bronchi	Thickness Gland mm.	Thickness Wall mm.	Ratio	No. of acini
Normal	7	8	·17	·64	·26	24
Chronic bronchitis	20	26	·64	1 07	·59	14
Bronchiectasis	7	10	·52	·93	·54	16
Emphysema without sputum	3	4	·23	68	·3	24

Table 1 shows the summarized results of these measurements from bronchi of seven normal subjects, of twenty patients with chronic bronchitis and of seven with bronchiectasis. The measurements were made in units of the micrometer eyepiece, but these have been converted to millimetres. Twenty-five counts were made on the eight normal bronchi and the average gland layer thickness found to be 0·17 mm., the maximum thickness 0·26 mm. The thickness of the wall internal to the cartilage was 0·64 mm., the ratio of gland to wall being 0·26. Sixty-five counts were made on 26 bronchi from 20 cases of chronic bronchitis (predominantly autopsy); the average gland-layer thickness was 0·64 mm. or three times the normal thickness, the corresponding depth of epithelium to cartilage being 1·07 mm., the increase in gland and wall thickness being 0 47 and 0·43 mm. respectively. Individual cases similarly show that it is the increase in the gland layer which is largely responsible for the overall increase in the wall. The gland forms less than one-third of the wall in the normal lung and about two-thirds in that of the chronic bronchitic. The bronchi were taken from the region of the hilum without any regard for the severity of the disease in the area they drained or for macroscopic appearance, and include main, lobar and segmental bronchi. As the

in the epithelium and relatively little in the glands. The large amount of secretion could represent the rapid transudation of fluid through these cells. The patient responded to prednisolone, is still on the drug, and recently cycled 70 miles in an afternoon. Biopsy after twelve months' control by prednisolone showed a respiratory-type epithelium, reduction in the congested submucus layer and a more normal appearance of the glands.

However, it cannot be assumed that the pathological changes seen in this patient are always present, particularly as in two cases to my knowledge prednisolone has failed to control the symptoms. Dr. F. H. Scadding has, moreover, had another case who responded to large doses of potassium iodide, while one reported case (Hartley and Davies, 1923) recovered with supportive measures only.

Hypersecretion of Mucus and Ciliary Activity

In many bronchitics the formation of mucus occurs abnormally close to the alveoli and in abnormally large amounts, both of which factors may strain the normal mechanisms for removal and may cause an accumulation of mucus in bronchi, in the tiny bronchioles or even in alveoli. The cilia may be overwhelmed by the secretion or it may be that the effective ciliary surface is reduced by the increased production of mucus, causing more goblet cells to discharge. Infection may destroy the epithelium, resulting in loss of ciliary function and effect. With healing, a layer of stratified epithelium is temporarily created, within which after about fourteen days there is differentiation of the cells to ciliated and mucus-secreting cells. This has been well demonstrated experimentally by Wilhelm (1953) and confirmed in man by Hers (1955).

Metaplasia to a stratified epithelium may be seen in patients with chronic bronchitis, but it is my experience that it is not a major feature. So also in bronchiectasis, it is not common in the bronchial tree except at the blind end of the dilated and obliterated bronchi, where a stratified epithelium often lines the residual sacs whose wall consists of fibrous tissue.

from consecutive operation specimens, mostly from patients with carcinoma. Dividing these into seven sputum groups—nil, trace, up to ℥ ss, ℥ ss to 1, 1 to 1½, 1½ to 2, and more than 2, the trend emerges that in each the ratio of gland to wall increases (Reid, 1960).

Surface epithelium. The epithelium in chronic bronchitis also shows an increase in goblet cells, most dramatically in the bronchioles, as normally only an occasional goblet cell is seen here (Plate XI, Fig. 2), while in chronic bronchitis (Plate XI, Fig. 4) almost every cell may be transformed to one distended with mucus. Even the 1 in 4 concentration of goblet to epithelial cells in the normal main bronchus is then surpassed in the periphery. Thus hypertrophy of mucous-secreting cells is present in bronchi, and also bronchioles and even the terminal bronchiole. These changes commonly occur throughout the lung, but they are irregularly distributed in that, in any one microscopic section, several square centimetres in area, some bronchioles may show change of the same degree, while others show very few cells. Mucus is thus produced much nearer the alveoli than in the normal; it is sometimes found free in alveoli and may be seen also in scars.

I was recently able to contrast the characteristic change in chronic bronchitis with the biopsy findings in a rare condition, pituitous catarrh, otherwise known as bronchitis serosa or bronchorrhea. A man of 70, a patient of Dr. F. H. Scadding, was incapacitated by the production of nearly a pint of phlegm a day, much of it looking like diluted egg white; the patient was breathless at rest and could walk only a few yards. Biopsy was performed at intervals of a week from two different sites in the right main bronchus and both showed a similar picture (Plate XII, Fig. 8). It is seen that the sub-mucosal layer of the bronchus is thickened; while the glands are to some extent responsible, this thickening is mainly between the glands and the surface in a layer of oedematous connective tissue with numerous capillaries. The epithelium is replaced by a transitional epithelium with no

Plate XI

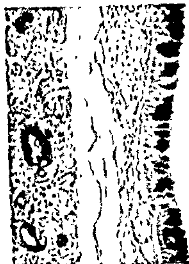


Fig. 1



Fig. 2



Fig. 3a

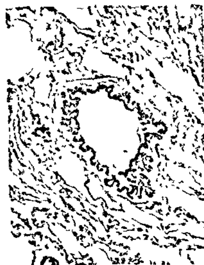


Fig. 4



Fig. 3b

Figs. 1–4. Periodic-acid-Schiff stain—mucus red. (1) Normal epithelium, [illegible]

(4) Bronchiole from subject dying with chronic bronchitis, showing increase in mucus-secreting cells (× 26)

Infection in Chronic Bronchitis

With regard to infection I wish simply to mention a few points relating to the histological appearances in chronic bronchitis. Plate XI, Figures 1–4 illustrate the changes which accompany hypersecretion of mucus. Both gland and surface epithelium may be in a state of hyperactivity without infiltration with inflammatory cells. Thus, histologically, hypersecretion is not necessarily associated with current infection, though infection may have been present at some stage.

The accepted picture of the damage caused by bacteria is the characteristic one of acute bronchiolitis, with its sequelae—stenosis or obliteration of bronchioles and, in alveoli, scarring and destruction giving rise to emphysema (Reid, L., 1958, 1959). This peripheral damage is too big a subject to discuss here but, to put it in context, it is this which is probably the most critical factor for disability or survival. However, if the role of bacteria is to be fully understood, lesser degrees and different types of infective damage must be taken into account.

Pus cells are often found in the sputum, but their presence does not necessarily mean ulceration of the surface epithelium, for they can migrate through the bronchiolar wall. This is perhaps better understood with eosinophils than with the neutrophils, though these latter can also migrate through epithelium as seen in the section in Plate XIII, Figure 9. Pus cells may be numerous in the lumen of a bronchiole whose epithelium is still intact, by reason of migration between the epithelial cells of polymorphonuclear leucocytes in response to bacteria present on the surface (Florey, 1933). Hers and Mulder (1953) have demonstrated migration by *H. influenzae.*

There are probably two levels at which bacteria may act—primarily on the surface where, as an irritant, they may cause secretion or migration of pus cells, and within the wall of the bronchiole. Inflammatory responses of both types are usually *irregularly distributed, even in autopsy specimens. For instance,* the main bronchi are often free of cellular infiltration, while peripherally there may be evidence of bronchiolar infection, either pus cells within the lumen or heavy infiltration within the wall, whether the epithelium is intact or damaged.

PLATE XI

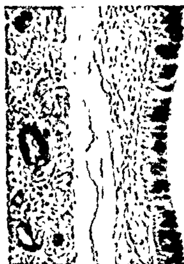


FIG 1



FIG 2



FIG 3*a*



FIG 3*b*



FIG 4

FIGS 1–4 Periodic-acid-Schiff stain—mucus red (1) Normal epithelium, [illegible]

mucus-secreting cells (× 26)

INFECTION IN CHRONIC BRONCHITIS

With regard to infection I wish simply to mention a few points relating to the histological appearances in chronic bronchitis. Plate XI, Figures 1–4 illustrate the changes which accompany hypersecretion of mucus. Both gland and surface epithelium may be in a state of hyperactivity without infiltration with inflammatory cells. Thus, histologically, hypersecretion is not necessarily associated with current infection, though infection may have been present at some stage.

The accepted picture of the damage caused by bacteria is the characteristic one of acute bronchiolitis, with its sequelae—stenosis or obliteration of bronchioles and, in alveoli, scarring and destruction giving rise to emphysema (Reid, L., 1958, 1959). This peripheral damage is too big a subject to discuss here but, to put it in context, it is this which is probably the most critical factor for disability or survival. However, if the role of bacteria is to be fully understood, lesser degrees and different types of infective damage must be taken into account.

Pus cells are often found in the sputum, but their presence does not necessarily mean ulceration of the surface epithelium, for they can migrate through the bronchiolar wall. This is perhaps better understood with eosinophils than with the neutrophils, though these latter can also migrate through epithelium as seen in the section in Plate XIII, Figure 9. Pus cells may be numerous in the lumen of a bronchiole whose epithelium is still intact, by reason of migration between the epithelial cells of polymorphonuclear leucocytes in response to bacteria present on the surface (Florey, 1933). Hers and Mulder (1953) have demonstrated migration by *H. influenzae*.

There are probably two levels at which bacteria may act—primarily on the surface where, as an irritant, they may cause secretion or migration of pus cells, and within the wall of the bronchiole. Inflammatory responses of both types are usually irregularly distributed, even in autopsy specimens. For instance, the main bronchi are often free of cellular infiltration, while peripherally there may be evidence of bronchiolar infection, either pus cells within the lumen or heavy infiltration within the wall, whether the epithelium is intact or damaged.

PLATE XIII

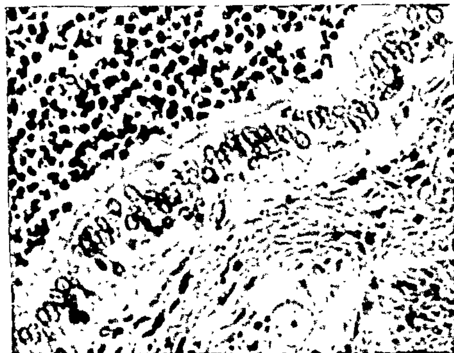


FIG 9 Part of bronchiolar wall plugged with pus cells, showing two leucocytes migrating between epithelial cells (Haematoxylin-eosin, × 280)

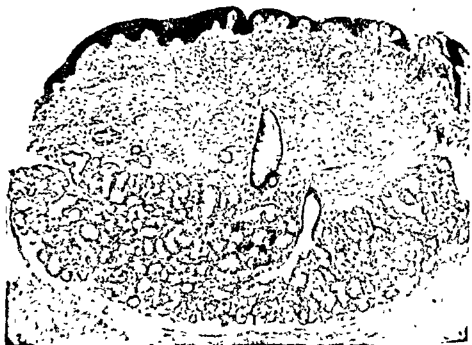


FIG. 8 Biopsy of bronchus from patient with pituitous catarrh, surface epithelium resembling a transitional epithelium, no goblet cells, sub-mucous layer vascular and increased in thickness. No increase in cells distended with secretion as would be seen in a patient with chronic bronchitis. (Haematoxylin-eosin, ×35.)

The exacerbations which characterize the natural history of chronic bronchitis are usually regarded as infective, because pus is often found, with or without pathogenic organisms in the sputum. There are, however, features which do not support this oversimplified picture—the frequent absence of pus cells, fever, antibodies in the blood corresponding to the 'pathogenic' organisms cultured from the sputum and the fact that pathogens cannot always be identified. The role of bacteria cannot be assessed simply as a recurrent invasion; they may assert themselves in a variety of ways—as a surface irritant causing an increased flow of mucus, by chemotactic effect giving rise to the passage of polymorphs through an intact epithelium, with or without an accumulation of polymorphs throughout the sub-mucosa. Even the 'non-pathogens' might, as an irritant, act as a damaging agent if they were sufficiently numerous.

With the excess of mucus which characterizes chronic bronchitis organisms commonly populate the bronchial tree, although normally below the larynx it is free of them The bronchial tree then shows some resemblance to the gut, in which a mucus-secreting surface constantly carries a bacterial population.

The suffix 'itis', of course, means inflammation, but not necessarily inflammation due to infection. Although the presence of pus cells in chronic bronchitis points to infection and Gram stains show the presence of bacteria, hypersecretion of mucus is also a sign of inflammation, as mucus-secreting surfaces respond to damage by secretion—a so-called 'catarrhal inflammation'. The word 'bronchitis' relates to inflammation of bronchi in this broad sense and is not restricted to infection.

RELATION OF CHANGES TO EPIDEMIOLOGICAL FINDINGS

What is it that causes hypersecretion? [illegible]

incidence of the disease or mortality.

A variety of conditions can be shown to be associated with a high incidence of chronic bronchitis; they range from housing conditions (Ogilvie and Newell, 1957) and atmospheric pollution (Logan, 1949, Pemberton and Goldberg, 1954; Stocks,

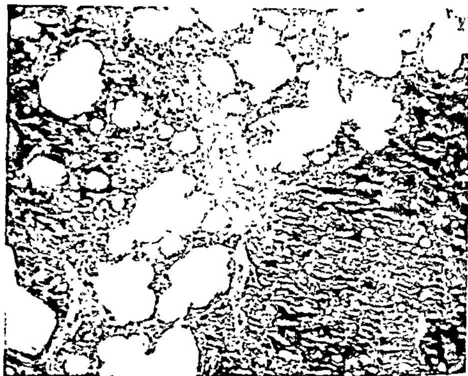


FIG. 10. Impaction of mucus (black) in alveoli of animal after five weeks' exposure to sulphur dioxide. Cultures from fives sites sterile. (Periodic-acid-Schiff, × 25)

anatomy of mucus-secreting structures. The first task was to establish the number of glands in several small laboratory animals. The hamster has only rudimentary glands in the trachea, each gland a dimple of epithelium with a few goblet cells and capillaries adjacent. Bronchi within the lung have very few goblet cells. In the mouse only an occasional goblet cell is present either in the trachea or along airways within the lung. Of larger animals the rabbit has very few mucus-secreting cells. The cat, on the other hand, has a large number of goblet cells and glands, which extend much further into the lung than in man. Strangely, however, a Periodic-acid-Schiff stain shows that the surface cells contain much mucin, but the glands in the wall comparatively little. This is a very different picture from that in the human lung and yet the behaviour of this species had to be drawn on in the attempt to understand human function.

The rat was the animal finally chosen. It is small compared with the hamster, but by contrast has well-developed glands in the trachea. They are multiple, branched tubular, and more numerous in the upper two-thirds and the anterior wall of the trachea. In the normal they are usually empty. Goblet cells are present in the epithelium of the trachea, numerous in main bronchi and along the main bronchial pathways within the lung. The finest bronchioles, and in particular the terminal, are free of them.

It was felt, therefore, that the rat offered the best chance of success in the experiments, even though it is prone to pulmonary disease, 'rat bronchiectasis' being a frequent, almost an invariable; finding in animals over eighteen months old; using the young rat we managed to exploit its propensity for secreting mucus while avoiding in the control animals the development of spontaneous infection.

These animals have been exposed to sulphur dioxide fumes for five days a week. A garden cloche mounted on a table was used, into which, through a hole in the base, were introduced three sampling tubes whose position could be varied and samples taken during exposure. It was found that the mixing of the fumes with air could be maintained at a constant level.

1959) to infection (Reid, D. D. and Fairbairn, 1958), smoking (Oswald and Medvei, 1955) and occupational factors (Lane, 1954; Reid, D. D., 1958). To attempt a synthesis between these findings and those of morbid anatomy it may be that all these stimuli act similarly by inducing hypersecretion of mucus. We know that in chronic bronchitis the amount of secretion is above the threshold level of sputum production, but how much above we do not know. With persistent irritation the tissues will suffer a work-hypertrophy—the factors which operate in any individual patient, and their relative importance, will vary.

The definition of chronic bronchitis adopted in various epidemiological studies has a twofold importance, first, that comparison of results must be made in relation to it and second, that one's attitude and further study of the disease may be influenced by it. If the definition includes disabilities such as shortness of breath ('Triple complex', Stuart-Harris, 1954; Oswald, 1958), incidents of infection and absence from work (Higgins *et al.*, 1956), it will take in more advanced stages than if cough and sputum (Ogilvie and Newell, 1957; Stuart-Harris, 1954) are the qualifications. If the earlier phases of the condition are to be studied—at a stage before medical treatment is sought or is at present possible—the latter definition may prove itself.

The main feature, then, of chronic bronchitis is *hypersecretion of mucus, but the presence of bacteria in the lung* is a further crucial element. Therefore, the questions which have to be answered are: (1) can irritation of itself lead to hypersecretion or must infection be present? and (2) if hypersecretion is itself a response to irritation, what role does infection then play? I have undertaken certain studies to test these questions experimentally.

Experimental Studies

By reason of its interest in relation to atmospheric pollution, sulphur dioxide has been chosen as the irritant in the experimental studies. Because of the key position which mucus-secretion holds in chronic bronchitis, in trying to produce the 'disease' in an animal, the aim has been to produce an animal which is hypersecreting mucus (Reid, L., 1960a).

The choice of animal presented a problem in the comparative

increased and most are distended with mucus, whereas in the normal many contain only granules. But what is especially notable is that the goblet cells develop in the finest airways in a part of the bronchial tree which is usually free of them. Mucus is not only found in the trachea but also in the bronchi and later it is found retained and impacted in the lung. Plate XIV, Figure 10 shows mucus in the alveoli of an animal killed after five weeks' exposure.

To return to the question to which the experimental studies have been directed, the selected agent has produced its effect by irritation of goblet cells and not by infection. Under normal conditions mucus is rapidly and effectively removed from the lungs and the lungs are empty and sterile although bacteria must frequently be inhaled. A simple mechanical explanation is that with the increase of mucus in the lungs, arising probably from both an increased production and slower removal, bacteria populate the bronchial tree.

While the dose has proved effective in these experiments to produce a hypersecreting animal, the size of dose of sulphur dioxide makes these experiments of little value in assessing the role of this irritant in causing human bronchitis, since our dose was 300 p.p.m. and the highest recorded concentration on the roof at St. Bartholomew's Hospital during the winter of 1958–9 was 1 6 p.p.m. (Lawther).

In brief, the stigma of chronic bronchitis is an increase in mucus-secreting tissue of all types at all levels—glands in the trachea and bronchi and goblet cells in the epithelium. The glands increase in total volume, while goblet cells appear in parts of the bronchial tree in which they are not normally present.

It would seem that a variety of irritants, including infection, may contribute to this increase. Once established, the excessive mucus seems to predispose to bacteriological population of the bronchial tree, which may act either as an irritant on the surface of the bronchi or by invading the wall and alveoli. These changes also seriously implicate the respiratory tissue, giving rise to damage and the respiratory failure that characterizes the later stages of the disease.

Large doses cause death through oedema and haemorrhage (Pattle). Exposure to 300 p.p.m. one day a week for six weeks brought no change, but 300 p.p.m. for five hours a day five days a week produced hypersecretion of mucus. To avoid bronchial irritation the animals were killed by intraperitoneal pentothal at weekly intervals and the thorax opened under aseptic conditions. Cultures were taken from each of five sites, the trachea, right and left main bronchi and both lungs, using sterile instruments for each. The presence of mucus in the trachea and bronchi was noted. In control animals no mucus was found and the trachea, both main bronchi and lungs in all were sterile. After three weeks mucus was seen by the naked eye in the exposed animals, which increased and persisted throughout the experiment and did not immediately disappear when exposure ceased. In most the mucus was sterile; no bacteria could be grown (virus culture was not attempted), which points to the fact that an excess of mucus was not necessarily associated with organisms within the lungs or bronchial tree. However, in some animals organisms were recovered, generally *H. influenzae* or *H. parainfluenzae*. The animals which died also showed these same organisms, a greater degree of mucus production and the presence of oedema.

Even if its rate of production is normal, there may be an excess of mucus in the bronchi; this may be due either to reduction in ciliary activity or to aspiration from the nose. The latter was investigated by placing India ink in the nares of animals before exposure, no India ink was found in the lungs. Further, animals subjected to sulphur dioxide often have obviously blood-stained mucus on their nose, but in no case was it found in the trachea.

The presence of mucus could mean that ciliary activity was so reduced that mucus remained in the lungs. Dahlam (1956) has shown that sulphur dioxide in smaller doses than we used does lower the rate of ciliary activity, but that recovery occurs if the cilia are not destroyed. The importance of this factor would be difficult to assess but, fortunately, there is histological evidence of increased cellular secretory activity in that goblet cells are also increased. The glands in the wall contain mucus; the number of mucous cells in the main bronchus epithelium is

LEONARDELLI, G B, PIGNATARO, O and VAGO, A. (1957). *Les Bronches*, 7, 580.
[illegible]

OSWALD, N. C. (1958). *Recent Trends in Chronic Bronchitis*. Lloyd-Luke, London.
OSWALD, N C. and MEDVEI, V C. (1955) *Lancet*, ii, 843
PATTLE, R. E Personal communication
[illegible]

REID, D. D (1958). *Lancet*, i, 1237 and 1289
REID, D. D and FAIRBAIRN, A. S. (1958). *Lancet*, i, 1147.
REID, L (1954). *Lancet*, i, 275
REID, L (1958). 'Pathology of Chronic Bronchitis and Emphysema' In *Recent Trends in Chronic Bronchitis*, ed. N. C. Oswald Lloyd-Luke, London.
REID, L (1959). *Brit J. Radiol.* 32, 291, 294
REID, L (1960) *Thorax*, (15, 132).
REID, L. (1960a) In press.
RHODIN, J (1957). *Les Bronches*, 7, 14.
STOCKS, P. (1959). *Brit med J* i, 74
[illegible]

WILHELM, D L. (1953) *J Path. Bact* 65, 543
WOOD, P. (1954). *Brit. med J* i, 1051.

I conclude, as I began, by saying that when we understand the origin and mechanics of hypersecretion we may be nearer the control of chronic bronchitis; we shall not solve the problem without that understanding.

ACKNOWLEDGEMENTS

The experimental studies described have been made possible by a grant from the Medical Research Council; Dr. P. Lawther and Mr. Cummins of the Atmospheric Pollution Unit helped in designing and setting up the exposure chamber; facilities for culture of animal lungs were given by Dr. R. W. Riddell in the Bacteriology Department of the Brompton Hospital. I am indebted to Dr. F. H. Scadding for permission to report his two cases of pituitous catarrh. The photomicrographs were prepared by Mr. D. F. Kemp of Kodak Limited.

REFERENCES

BABKIN, B. P. (1950) *Secretory Mechanism of the Digestive Glands.* 2nd ed., Paul B Hoeber, Inc., New York.
BINET, L. and BOURLIÈRE, F. (1944). *Sem Hôp , Paris,* 20, 284.
BUCHER, U. G. and REID, L. (1960). In press.
DALHAMN, T. (1956). 'Mucous Flow and Ciliary Activity in the Trachea of Healthy Rats and Rats Exposed to Respiratory Irritant Gases.' *Acta phys. scand.* 36, supplement 123.
FLOREY, H (1930). *Brit J exp Path.* 11, 348.
FLOREY, H. W. (1933). *J. Path Bact* 37, 283.
FLOREY, H., CARLETON, H. M and WELLS, A. Q (1932). *Brit. J. exp Path.* 13, 269.
FORBES, J. and WISE, L (1957) *Lancet,* ii, 767.
GAMMARROTA, V. (1954). *Riv Tuberc App. Resp.* 2, 125.
HARTLEY, P. H -S. and DAVIES, I J (1923). *Brit med. J.* 1, 1052.
VON HAYEK, H (1953) *Die Menschliche Lunge Springer* Berlin
HERS, J. F. PH. (1955). *The Histopathology of the Respiratory Tract in Human Influenza* Leiden, H E. Stenfert Kroese.
HERS, J. F PH. and MULDER, J. (1953). *J Path Bact.* 66, 103.
HIGGINS, I T. T , OLDHAM, P. D , COCHRANE, A L. and GILSON, J. C. (1956) *Brit. med. J* ii, 904.
HILDING, A. C. (1943) *Ann Otol. Rhinol. Laryngol.* 52, 5.
HOUSTON, J. C , DE NAVASQUEZ, S and TROUNCE, J R. (1953). *Thorax,* 8, 207
LANE, R E. (1954). *Trans. Ass. industr. med. Offrs* 4, 58.
LAWTHER, P. Personal communication.

lying cause can be removed [illegible]
noses malignancy on the a[illegible]
pathologist by the presence i[illegible]
necrosis of terminal arteries and arterioles and proliferative intimal thickening of slightly larger vessels. Yet neither has any clear picture of what lies behind this dramatic transformation. The significance of hypertensive encephalopathy is that it enables us to bridge the gap in knowledge.

The various causes of hypertension all seem to act by causing an increase in the tone of the systemic arterioles, to which the heart responds by contracting more strongly and so raising the systemic blood pressure. It will be observed that the plain muscle fibres in the arterioles are consequently between two fires, for they are being at once overstimulated to contract and overstretched by the resulting rise in filling tension. The most remarkable fact about hypertension is that in the great majority of cases the arterioles contrive to the end, despite this double strain, to serve their normal function of accurately controlling the distribution of the circulating blood. So it is that in the benign phase serious symptoms are deferred until the heart fails or a vital artery bursts or becomes blocked by secondary atheroma. But there is a limit to the adaptability of the arterial muscle and it is when this limit is exceeded that the essential circulatory disturbance of malignancy occurs. I propose first to state the nature of this disturbance as I see it and then to outline the evidence on which I base this statement.

The sequence of events seems to be as follows. When the underlying vasoconstrictor stimulus, whatever its nature, becomes too intense or develops too rapidly, the excessive intra-arterial tension directly provokes a state of cramp-like contracture, or spasm, in scattered groups of arterioles and terminal arteries, which in turn leads to focal anoxia, anaemia or asphyxia of the tissue supplied. The effects of this tiny vascular crisis vary with its severity, its duration and its extent. Thus an acute widespread crisis will cause anoxia of whole organs and *since the brain is most sensitive to anoxia this will commonly* cause an attack of acute encephalopathy, of which the eclamptic convulsion is a typical example. Such an attack may be too

XV

The Significance of Hypertensive Encephalopathy

F. B. BYROM

When the blood pressure in hypertensive disease becomes unusually high or rises very steeply two apparently unrelated complications are apt to occur. The first is the acute disturbance of cerebral function known as hypertensive encephalopathy. The second is a crop of acute necrotic lesions involving terminal arteries and arterioles in many tissues—the lesions of malignant hypertension. In the course of a long-term study of experimental hypertension it became increasingly clear that the two were different expressions of a single focal but widespread vascular disorder, the course of which could be followed in some detail in the brain of the hypertensive rat.

In reporting these findings in detail (Byrom, 1954) I was concerned primarily with the problem of encephalopathy and only incidentally with malignant hypertension. In the present lecture I shall reverse the emphasis and try to show how a study of experimental encephalopathy throws direct light on the early stages of the local vascular catastrophe which underlies malignancy.

It is now generally accepted that hypertensive disease, although deriving from a number of different causes, known and unknown, can be divided into two sharply contrasted phases or grades of severity, benign and malignant, the former usually persisting for years without causing serious symptoms, the latter spelling death within months or even weeks unless the under-

malignancy were found in many organs but not in the kidney distal to the clamp. From this observation, which has been repeatedly confirmed, it follows that the lesion cannot be caused by a circulating toxin, by simple ischaemia, nor by simple vasoconstriction by a circulating pressor substance, but must in some way be a result of excessive intra-arterial tension. It seemed that the determining factor was a sudden strain imposed on the vessel wall by the combination of severe vasoconstriction and the excessive filling tension. A similar distribution of lesions was observed by Goldblatt in the dog but since he observed lesions only when the constriction of the renal artery was severe enough to cause renal insufficiency he believes that this, as well as hypertension, is essential to their causation (Goldblatt, 1947). To produce severe and persistent hypertension in the rabbit or the dog it is necessary to constrict both renal arteries or one artery after excising the opposite kidney and severe constriction in such circumstances is, of course, liable to be complicated by renal insufficiency. Fortunately it was soon found that in the rat permanent hypertension could be caused by constricting one renal artery, the opposite kidney being left intact (Wilson and Byrom, 1939). In such animals typical arterial lesions appeared in the usual sites and were absent from the clamped kidney. The opposite kidney, however, displayed not only arterial necrosis, but also and in minute detail the characteristic changes of malignant hypertension. This experiment demonstrated clearly that renal insufficiency was not necessary to the production of the arterial lesions or, in other words, that the difference between benign and malignant hypertension is simply one of degree. This simple concept of malignancy, which does not necessarily imply that the critical level of pressure is the same in all cases, is supported by the demonstration that arterial necrosis can be caused by sudden distension of the arterial tree, in animals by forcible injection of saline (Byrom and Dodson, 1948; Masson, Corcoran and Page, 1951; Magarey, 1956) and in man by operative relief of coarctation of the aorta (Benson and Sealy, 1956). Necrosis is also found in the pulmonary arteries in mitral stenosis. Furthermore it is now well recognized that some forms of unilateral renal disease in man may cause

short to damage tissue. On the other hand less widespread but more lasting crises recurring at short or even imperceptible intervals will lead to a succession of crops of scattered small lesions. These may be of three grades, viz. (1) mild and reversible, consisting of small zones of increased capillary permeability and focal oedema, (2) severe enough to kill tissue, that is, to cause small infarcts or haemorrhages, or (3) severe enough to kill the artery itself.

Once this pathological process has become firmly established the patient's life is in danger because the process leads to progressive renal damage of a kind which not only leads to uraemia, but also reinforces the original hypertension (Wilson and Byrom, 1941). I may add that I have long held the view (Byrom and Dodson, 1949) that hypertensive disease is irreversible only in so far as it may derive from, or has itself caused irreparable damage involving both kidneys—and nowadays even this obstacle can be surmounted, at least in twins, by kidney grafts (Murray, Merrill and Harrison, 1958).

The evidence from which the above picture is derived is as follows.

The Arterial Lesions of Malignant Hypertension are Results of Excessive Intra-Arterial Tension

The fact that almost any kind of hypertension can become malignant indicates that malignancy is not determined by any of the underlying causes of the disease and suggests that it is related in some more direct way to the hypertensive process itself. Post-mortem studies have shown that arterial necrosis, proliferative endarteritis and associated tissue damage are the essential lesion of malignant hypertension and are only occasionally seen in the benign phase. But morbid anatomy throws no light on the cause of this lesion and for many years it was regarded as 'inflammatory' or 'allergic' in nature. The true nature of the lesion became quite clear soon after Goldblatt's work opened the field of hypertension to experimental study. The crucial piece of evidence was the observation of Wilson and Pickering (1938) that in rabbits with severe hypertension caused by constricting the artery to a solitary kidney typical lesions of

malignancy were found in many organs but not in the kidney distal to the clamp. From this observation, which has been repeatedly confirmed, it follows that the lesion cannot be caused by a circulating toxin, by simple ischaemia, nor by simple vasoconstriction by a circulating pressor substance, but must in some way be a result of excessive intra-arterial tension. It seemed that the determining factor was a sudden strain imposed on the vessel wall by the combination of severe vasoconstriction and the excessive filling tension. A similar distribution of lesions was observed by Goldblatt in the dog but since he observed lesions only when the constriction of the renal artery was severe enough to cause renal insufficiency he believes that this, as well as hypertension, is essential to their causation (Goldblatt, 1947). To produce severe and persistent hypertension in the rabbit or the dog it is necessary to constrict both renal arteries or one artery after excising the opposite kidney and severe constriction in such circumstances is, of course, liable to be complicated by renal insufficiency. Fortunately it was soon found that in the

in the usual sites and were absent from the clamped kidney. The opposite kidney, however, displayed not only arterial

tion of the arterial lesions or, in other words, that the difference between benign and malignant hypertension is simply one of degree. This simple concept of malignancy, which does not necessarily imply that the critical level of pressure is the same in all cases, is supported by the demonstration that arterial necrosis can be caused by sudden distension of the arterial tree, in animals by forcible injection of saline (Byrom and Dodson, 1948; Masson, Corcoran and Page, 1951; Magarey, 1956) and in man by operative relief of coarctation of the aorta (Benson and Sealy, 1956). Necrosis is also found in the pulmonary arteries in mitral stenosis. Furthermore it is now well recognized that some forms of unilateral renal disease in man may cause

benign or malignant hypertension which responds dramatically to nephrectomy provided that this is done before significant secondary hypertensive damage has developed in the opposite kidney. Finally it is widely accepted that renal insufficiency is often absent in the early stages of malignant essential hypertension (Ellis, 1938).

All these facts support the simple quantitative concept of malignancy and it is probably true to say that it is now generally accepted. Goldblatt's (1947) opinion that hypertensive arterial necrosis occurs only in the presence of renal insufficiency is not consistent with the evidence described above and his view that the renal lesions in the hypertensive rat were due to spontaneous pyelonephritis (Goldblatt, 1947) or polyarteritis nodosa (Goldblatt, 1958) ignores the photographic evidence (Wilson and Byrom, 1939, 1941) and the fact that in these experiments the lesions were invariably absent from the clamped kidney. More recently the current hypothesis has been questioned by Kincaid-Smith, McMichael and Murphy (1958). These authors do not deny that hypertension can cause arterial necrosis but they consider it unwise to assume that this is the only type of injury concerned. They suggest that the appearances could be the same if the vessels were 'damaged by some toxic influence or even by allergic injury' and use the term 'exudative angeitis' to cover the various lesions. In the early years of the present century malignant hypertension was completely confused with primary nephritis and the recognition of its separate nature has been generally accepted as an important step towards the clarification of one of the most obscure problems in medicine. To postulate unspecified toxins or allergens must necessarily revive old confusions and the Hammersmith workers' reasons must therefore be closely scrutinized. Their first point, that the vascular lesions of malignancy closely resemble those of allergy, is not disputed, but, as I have pointed out elsewhere (Byrom, 1954), a common structural lesion in disease does not necessarily imply a common mechanism, much less a common cause, for a tissue has only a very limited number of ways of reacting against an almost infinite range of injuries. Two further arguments advanced by these authors are that although the blood pressure

tends to be very high in the malignant phase it is not impressively higher than in patients with severe benign hypertension and retinitis, and that scattered arterial necrosis is not uncommonly found post-mortem in benign hypertension. In my opinion both these facts are consistent with the present hypothesis. For if the difference between the two phases is simply one of degree the blood pressure in severe benign hypertension must closely approach the malignant level and even exceed it if due allowance is made for relevant factors such as individual variability and the steepness of the rise of pressure. By the same token arterial lesions are to be expected occasionally in the benign phase because it takes more than one vascular crisis to constitute malignancy. A final argument is that in the kidney in malignant hypertension it is common to find arterial necrosis in afferent glomerular arterioles distal to intralobular arteries which are greatly narrowed by proliferative endarteritis, which should, on the mechanical hypothesis, protect the smaller vessels against necrosis. The weakness of this argument is that it presumes that the lesion in the larger vessel precedes that in the arteriole. There is no evidence that it does and elsewhere in their paper the authors accept the experimental evidence and adduce further clinical evidence that narrowing of a large renal artery protects the smaller vessels against both types of lesion. I have discussed this problem at length because the simple quantitative explanation of malignancy is crucial to the modern conception of hypertensive disease. The hypothesis must remain tentative as long as so little is known about aetiology. But I believe it provides the simplest explanation of the known facts and it is a sound principle in science to avoid multiplying hypotheses unnecessarily.

The Production of Hypertensive Encephalopathy in the Rat

To pursue the problem further it is necessary to turn to a small animal, preferably the rat. One reason for this is that although the arteries differ widely in size in man and the rat the arterioles and capillaries do not, so that a focal vascular upset at the arteriolar level which would be a trivial matter in man may in the rat be a major disaster which by causing gross cerebral

symptoms would regularly attract attention in the early stages A further advantage is that the brain is so small that the presence and number of organic lesions can easily be determined by a few serial sections at intervals of about 0·5 mm. I have therefore used young adult male rats made hypertensive by constricting the left renal artery with a silver clip after removing the right kidney. Rats so treated regularly develop severe hypertension which is at any time liable to be complicated by acute attacks of general or local convulsions, ataxia, weakness and coma, that is by typical acute encephalopathy. Apart from occasional rats which are moribund at the time of operation and remain in coma, these attacks can be abruptly terminated by removing the clip from the renal artery, a procedure which always abolishes the hypertension in a few hours. The hypertensive origin of the symptoms is therefore not in doubt. In the rat, moreover, as in man it has been shown that renal insufficiency plays no part in causing the symptoms. With this preparation it becomes possible to make serial observations on the same rat before cerebral symptoms appear—that is, in the stage of *simple* hypertension—during cerebral attacks, and finally after abolishing both the hypertension and the symptoms.

ENCEPHALOPATHY AND MALIGNANCY ARE RELATED PHENOMENA

The symptoms just mentioned are clearly acute attacks of hypertensive encephalopathy. What evidence is there that they are *related to the equally acute essential lesions of malignancy?* This question was readily answered by examination of serial sections of the brain. In fourteen rats with severe simple hypertension two healed infarcts were discovered but no acute lesions were found. On the other hand, in 196 rats with encephalopathy organic lesions were found in 108 (55 per cent). The lesions were focal, usually multiple and varied in appearance. Acute arterial necrosis or proliferative endarteritis occurred in 26 per cent, recent miliary infarcts in 21 per cent, capillary haemorrhages in 30 per cent, larger haemorrhages in 22 per cent and healed infarcts in 12 per cent. These are all components of the essential lesion of malignant hypertension and they clearly occur commonly enough in encephalopathy to indicate that the two

phenomena are in some way related. It can be stated definitely that the structural lesions are not the cause of the cerebral symptoms, (1) because lesions were absent in 45 per cent of attacks of encephalopathy and (2) because when the clip was removed from the renal artery convulsions ceased immediately and other symptoms within a few hours, that is, long before organic lesions could heal. The only reasonable explanation of the association is that the symptoms are an early and lesions a later and by no means inevitable result of a single functional disturbance precipitated by the high blood pressure, a disturbance which is manifestly vascular in nature and focal, though widespread, in distribution.

Before considering the nature of this primary disturbance it is necessary to examine another of its effects which is of considerable interest and significance, namely focal oedema.

Encephalopathy and Malignancy are both Related to Focal Oedema

Oedema of the brain has occasionally been reported in fatal eclampsia, while in rats dying from encephalopathy a conspicuous cerebellar 'pressure cone' (Plate XV, Fig. 1)[1] is commonly discovered post-mortem. For these reasons the water content of the brain was measured after death in a large number of rats. In simple hypertension the readings were found to be normal, but in encephalopathy a significant increase of brain water occurred, slight at first, but more marked in the later stages.[2] What is the significance of this oedema? Is it a simple oedema due to an alteration in the effective filtration pressure in the capillaries? Or does the underlying vascular disturbance cause a focal increase in capillary permeability or in the affinity of the tissues for water? To study this question a convenient tool is provided by trypan blue, a dye which when injected intravenously fails to stain the normal brain because it cannot penetrate the blood-brain barrier, unless this barrier is injured. The dye is therefore a useful marker of acute capillary damage.

[1] The plates referred to in this lecture will be found between pages 264–5.
[2] For full data see Byrom, 1954.

A further series of rats was accordingly injected with trypan blue and killed 30 minutes later. In 45 rats with severe simple hypertension the brain remained unstained in all cases. But out of 145 rats with encephalopathy no fewer than 126 (87 per cent) *showed multiple blue spots on the surface of the cerebral cortex* post-mortem (Plate XVI, Fig. 2).

The next step was to compare the water content of the stained areas with that of the intervening brain. This study showed a clear increase of water in all brain tissue containing blue zones. The results also showed a less marked but definite oedema in the unstained regions of the brain in late but not in early cerebral attacks. Thus the cerebral oedema in encephalopathy is primarily a local phenomenon associated with increased capillary permeability, but if the condition is not relieved this soon becomes overlaid by a generalized oedema. I suspect that this last is a secondary result of venous obstruction, but whatever its cause it is of great practical importance because it involves enough brain tissue to cause death from respiratory paralysis. The earlier focal oedema is not on a large enough scale to be dangerous but it is of greater theoretical moment in two respects. First, it provides additional evidence to that already *mentioned of the regular recurrence* of a focal vascular disturbance in rats with cerebral symptoms. Secondly, it raises the question of what relation if any exists between these zones of focal oedema and the acute structural lesions described above. This question was answered by histological study of a large number of brains containing blue zones. This study showed that wherever an active organic lesion was observed it was found to tally with a blue area. A much more significant finding was that blue areas were five times as numerous as organic lesions. These *findings provide a further link in the chain joining encephalopathy* and malignant hypertension. We are clearly dealing with a focal vascular disturbance which usually (though not invariably) causes in the brain multiple areas of increased capillary permeability. Most of these areas show nothing more than local oedema but in a significant minority the process is sufficiently intense to cause the typical lesion of malignant hypertension. Amongst the most disturbing symptoms of this disease are

Plate XV



Fig 1 Post-mortem dissection of the back of the neck of a rat dying from encephalopathy. The occipito-atlantal membrane has been removed, showing a bulbous protrusion of the vermis of the cerebellum through the foramen magnum (× 5).



Fig 5 Spasm of arteries on the surface of the small intestine in a rat with encephalopathy. (× 10)

morning headache, vomiting and failing vision. These symptoms are accompanied by papilloedema and are clearly due to increased intracranial tension. The demonstration that the typical lesions of the disease are accompanied by much more numerous foci of oedema suggests a simple explanation of these symptoms, an explanation which implies that the sustained cerebral symptoms of malignant hypertension may properly be described by the term [illegible].

To what extent is f[illegible]ute cerebral symptoms in [illegible] is abolished in such rats by removing the clip from the renal artery, focal myoclonic contractions, when present, often persist for a few hours and may therefore be due to oedema. On the other hand generalized convulsions are always abolished immediately by removing the clip and must therefore be caused by some more rapidly reversible factor. Moreover focal oedema is not demonstrable, at any rate by vital staining, in 13 per cent of attacks.

Evidence for Arterial Spasm

It has long been suspected that cerebral arterial spasm is at the root of encephalopathy but direct evidence has been entirely lacking. Measurement of the carotid blood flow in the rat showed a progressive fall in encephalopathy, but this could well be a secondary result of cerebral oedema. The problem was therefore examined more directly by developing a technique (Byrom and Cameron, 1955) for inserting permanent windows of moulded acrylic resin in the dorsal surface of the skull which could be covered over with skin and exposed on suitable occasions for observation and photography of the cerebral arteries. The windows were inserted in rats with severe simple hypertension before any cerebral attacks had been observed. The results of this photographic study were as follows (Plate XVI, Figs. 3, 4). On 190 occasions (160 rats) in rats with simple hypertension the arteries almost always appeared normal, spasm being observed on only 7 occasions. In encephalopathy, on the other hand, diffuse and/or focal arterial spasm with pallor of the surface of the brain was recorded on 134 occasions (120 rats) and was absent on 16 occasions (15 rats). In 76 rats control

photographs were taken several days after removing the clip from the renal artery. The vessels always appeared normal.

In assessing these findings allowance must be made for the fact that the brain starts swelling very early in an attack of encephalopathy and in many photographs the vascular pattern during attacks is distorted and it may fairly be suggested that the spasm was secondary to such distortion. But even when all cases showing distortion are excluded there remain 20 instances (18 rats) in which well-marked arterial spasm was present and only 2 instances (2 rats) on which spasm was absent. I regard this as strong evidence that the spasm so regularly observed in cerebral attacks is genuine, but in such an important question there is clearly room for further study by other techniques.

Changes in Other Organs—the Concept of the 'Vascular Crisis'

The concept of the 'vascular crisis' is not new. According to Volhard (Berglund and Medes, 1935) J. Pal (1905) long ago recognized on purely clinical grounds that some hypertensive patients were subject to sudden attacks in which abrupt elevations of blood pressure were accompanied by symptoms referred to various organs and he termed these attacks 'vascular crises'. At that time encephalopathy was generally regarded as an expression of uraemia and malignant hypertension had not been distinguished from primary Bright's disease. If these two conditions are in fact different facets of the vascular crisis envisaged by Pal, the changes which I have described in the brain should be and in fact are demonstrable in other organs of rats displaying cerebral symptoms. Typical structural lesions were common in the intestine, pancreas and heart. Local oedema was commonly seen in the pancreas and occasionally in the wall of the intestine. Arterial spasm is regularly seen in the intestinal arteries (Plate XV, Fig. 5) and often in the retina (Plate XVII, Figs. 6, 7) and I have also observed lateral detachment of the retina (Plate XVII, Fig. 8), another example of focal oedema. But several fundamental questions remain unanswered. First, it is not clear why excessive intra-arterial tension should precipitate focal arterial spasm. It may be (Byrom, 1954) that it is an

exaggeration of a primitive tendency of plain muscle to contract against a stretching force (Bayliss, 1902) but as long as the cause of the underlying vasoconstriction in hypertension remains obscure the question seems likely to remain unanswered. Secondly, to what extent does the pulsatile component of the blood pressure, which is considerably increased in hypertension, contribute to the strain which precipitates spasm? Inspection of the tortuous pulsating arteries in the mesentery of the hypertensive rat will leave the observer with the strong impression that the contribution is a large one. In operating on such animals to remove a renal arterial clamp I have repeatedly been impressed by the virtual absence of pulsation in the renal artery distal to the clamp and by the striking increase which occurs as soon as the clamp is removed. Lastly, while it is easy to understand how spasm can cause cerebral symptoms, focal oedema and death of tissue, it is still not clear how it leads to local death of the artery itself. In recent unpublished observations, however, I have found that single intravenous injections of very large doses of synthetic hypertensin (Ciba) regularly cause renal arterial necrosis in circumstances which support the view that sudden rises of blood pressure cause necrosis simply and directly by overstretching the contracted muscle fibres of the media.

REFERENCES

BAYLISS, W M (1902) *J Physiol* 28, 220.

BENSON, W. R. and SEALY, W C (1956) *Labor Invest* 5, 360

BERGLUND, H and MEDES, G. (1935). *The Kidney in Health and Disease* Lea and Febiger, New York, p. 665.

BYROM, F B (1954) *Lancet*, ii, 201.

BYROM, F. B and CAMERON, D A (1955). *Austral. J exp Biol and Med Sci.* 33, 225

[illegible]

Med 27, 117.

MAGAREY, F. R. (1956). *Med. J. Austral.* Sept. 29, 473.
MASSON, G. M. C., CORCORAN, A. C. and PAGE, I. H. (1951). *Revue Canadienne de Biologie*, **10**, 309
MURRAY, J. E., MERRILL, J. P. and HARRISON, J. H. (1958). *Ann. Surg.* **148**, 343.
PAL, J. (1905). *Die Gefässkrisen* Hirzel. Leipzig.
[illegible]

XVI

Viability and Survival of Red Cells

T. A. J. PRANKERD

THE challenge of haemolytic anaemias to different investigators has appeared in different forms. To early workers the problem naturally appeared morphological, and emphasis was placed on variations in cell size and shape. Later as interest grew in the *in vitro* properties of blood the emphasis shifted to various chemical agents which haemolysed red cells, and later still with the advent of splenectomy it shifted to the spleen as a site of red cell destruction. Today, antibodies dominate the field, and the importance of the cell itself and its living activities are being appreciated. It is the purpose of this lecture to consider only the living aspects of the red cell, to examine these and try to assess their significances in the life of the red cell. By living processes are meant those that are in a state of dynamic equilibrium, and which depend for the upkeep of this equilibrium on energy processes within the cell.

Red Cell Metabolism

The red cell differs metabolically from other cells in certain particular features. There is no glycogen in the cell and glucose appears to be the principal substrate, being degraded anaerobically via the Embden Myerhoff pathway; the ultimate chemical process is the formation of lactate from pyruvate and its diffusion out of the cell. There is therefore no store of metabolite as exists in most cells. The enzymes of the tricarboxylic acid cycle are not present in complete sequence so that pyruvate and acetate cannot be oxidized. Two unusual features are the presence of large amounts of 2,3-diphosphoglycerate and of

glutathione; both compounds occur of course in other cells, but in them only in catalytic amounts. The possible reasons for the presence of these compounds in the red cell will be considered later, but an important role of 2,3-diphosphoglycerate

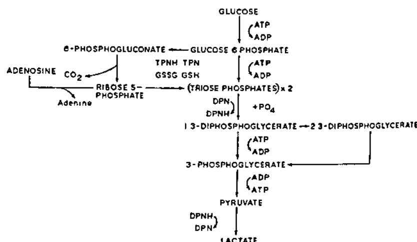


FIG 1 Scheme of some metabolic events in the red cell.

is that it can act as a store of phosphate and energy in the cell much as creatine phosphate does in muscle. It appears that the hexose monophosphate shunt operates, but in normal circumstances it is probably not very active, however certain reactions in this cycle are potentially important. These sequences of metabolic events are shown in Figure 1 and have been previously reviewed (Prankerd, 1956).

The degradation of glucose results in the formation of energy which is stored in the cell in chemical form as the high-energy bonds of purine nucleotides and 1,3-diphosphoglycerate. This energy is required by the cell for several purposes, but the most important function of the red cell, that of combining and transporting oxygen, is a simple physical process not dependent on any source of energy. The energy in the cell appears to be diverted to other functions which are those most intimately concerned in maintaining the survival of the cell itself in the circulation.

First of these functions and possibly most important is the

maintenance of the cation distribution peculiar to all cells. To maintain this electrochemical gradient, energy is required to extrude sodium from the cell and take up potassium. There are many theories about the mechanisms involved in this process, which will not be considered here, but this function can be calculated to involve about 7 per cent of the total energy formation of the red cell, and in the face of a breakdown of these mechanisms the cell becomes osmotically unstable since there is nothing to oppose the osmotic pressure exerted by the high intracellular concentration of haemoglobin; a cell, such as the red cell, which behaves as an osmometer must therefore progressively swell and burst, unless other changes intervene.

The second function to be considered is the maintenance of the integrity of the cell membrane. There is evidence suggesting that the lipid fraction of the membrane exists in a dynamic state although the protein does not do so. The evidence for this is not yet complete, but several workers (Altman, 1951; James, Lovelock and Webb, 1957) have shown that labelling of lipids occurs if blood is incubated with ^{14}C-acetate, and that the lipids so labelled are then released into the plasma. What is not certain is how much of the synthesis of lipid in blood is due to white cells. Buchanan (1960) has shown that the red cell contribution must be very small. Some small extent of synthesis may occur in red cells as lipids can be washed off the red cell membrane very readily and if these cells are then reincubated in plasma they appear to regain some of their lost lipid, a phenomenon which would probably only occur by synthesis. Whether we accept the existence of an energy-dependent dynamic state of the lipids of the red cell wall is therefore still uncertain, but the lipids themselves are of vital importance in maintaining the continuity of the cell membrane.

The next energy-dependent function to consider is the mechanism involved in maintaining glutathione in its reduced state. Oxidation of glutathione occurs rapidly in a cell deprived of glucose, but this reaction is reversed if glucose or a nucleoside is present; no other substrate is effective. A flavoprotein, glutathione reductase is the immediate enzyme involved in glutathione reduction, and it activates the donation of hydrogen from

reduced triphosphopyridine nucleotide which is itself reformed continuously by the oxidation of glucose to 6-phosphogluconate. The enzyme involved is glucose 6-phosphate dehydrogenase. It is not known why the red cell is so rich in glutathione although a possible reason is that it is required to protect haemoglobin from oxidation *in vivo* by such substances as hydrogen peroxide and various drugs; it also appears to be involved in maintaining lipid stability in the membrane. Lastly it should be mentioned that the peculiar biconcave shape of the red cell may also be dependent on a source of energy for its maintenance, but direct evidence is lacking on this point.

Experimental Factors Influencing Red Cell Survival

For many years, until perhaps the 1940s, it was usual to regard the red cell as inert and lifeless and to assume that its ultimate destruction was the result of a physical process attributable to wear and tear. However, red cell survival curves simulate those of any population diminishing by a process of ageing, and two experiments can be described to show that the 'wear and tear' hypothesis is not correct and that in considering mechanisms of red cell destruction in the body the metabolic state of the cell must be taken into consideration. The first of these experiments depends upon the changes occurring in red cells during storage. After about six weeks of storage at 4° C. in acid-citrate-dextrose media radical changes have occurred in the composition of red cells. There is a considerable breakdown in phosphate ester, particularly adenosine triphosphate (ATP) and 2,3-diphosphoglycerate. If these cells are now reincubated with glucose at physiological pH no utilization of sugar occurs and the metabolic state of the cell only deteriorates. Cells which have been stored in this way are more spherical than normal, contain more sodium and potassium and their membranes are deficient in lipid. When injected into the circulation of a compatible recipient they are rapidly removed, mainly by the spleen. These cells are thus non-metabolizing and non-viable. If, however, the cells are taken and incubated for half to one hour with a suitable nucleoside a remarkable change occurs in them. They resynthesize adenosine triphosphate, and 2,3-diphosphoglycerate,

and begin again to metabolize glucose, so that if they are now injected into circulation their survival is greatly improved. These two changes are shown in Table 1. The mechanism through which this revival is brought about is probably the following.

TABLE 1. Percentage survival + ATP contents of cells stored for varying periods and incubated with *nucleosides immediately* before transfusion

Days of storage	Nucleoside	% survival at 48 hrs.		ATP content of cells ugP/ml. cells		
		Nucleoside	Glucose	Before incubation	After incubation Nucleoside	After incubation Glucose
28	Guanosine	75	54	25	54	30
28	Adenosine	80	28	21	41	10
42	Guanosine	69	28	21	36	0
42	Adenosine	51	10	16	27	0
56	Guanosine	49	10	16	29	0

When a cell's ATP stores have fallen below a critical level there is no longer sufficient present to assist the initial phosphorylation of glucose by hexokinase and the cell is apparently dead. If, however, the other source of sugar, the nucleoside, is present, ribose is split from the purine by a nucleoside phosphorylase and converted to ribose 5-phosphate which can then be utilized in the monophosphate shunt with the formation of triose phosphate *and thence pyruvate*, and the resynthesis *of ATP; the reactions* are made clear in Figure 1. In this way ATP is built up again in the cell and can once more be utilized for the phosphorylation of glucose. The apparently dead cell is now viable. The con- [illegible]

[illegible] conclusions can be justified in another way, in this case by the use of metabolic inhibitors. Thus if blood is incubated with sodium arsenate for two hours the arsenate radicle replaces that of phosphate in esters within the cell; compounds of this sort can no longer be utilized and glycolysis stops. Cells treated in this way and labelled with an isotopic *tag* can be *injected* into the circulation, from which they rapidly disappear after

T

reduced triphosphopyridine nucleotide which is itself reformed continuously by the oxidation of glucose to 6-phosphogluconate. The enzyme involved is glucose 6-phosphate dehydrogenase. It is not known why the red cell is so rich in glutathione although a possible reason is that it is required to protect haemoglobin from oxidation *in vivo* by such substances as hydrogen peroxide and various drugs; it also appears to be involved in maintaining lipid stability in the membrane. Lastly it should be mentioned that the peculiar biconcave shape of the red cell may also be dependent on a source of energy for its maintenance, but direct evidence is lacking on this point.

Experimental Factors Influencing Red Cell Survival

For many years, until perhaps the 1940s, it was usual to regard the red cell as inert and lifeless and to assume that its ultimate destruction was the result of a physical process attributable to wear and tear. However, red cell survival curves simulate those of any population diminishing by a process of ageing, and two experiments can be described to show that the 'wear and tear' hypothesis is not correct and that in considering mechanisms of red cell destruction in the body the metabolic state of the cell must be taken into consideration. The first of these experiments depends upon the changes occurring in red cells during storage. After about six weeks of storage at 4° C. in acid-citrate-dextrose media radical changes have occurred in the composition of red cells. There is a considerable breakdown in phosphate ester, particularly adenosine triphosphate (ATP) and 2,3-diphosphoglycerate. If these cells are now reincubated with glucose at physiological pH no utilization of sugar occurs and the metabolic state of the cell only deteriorates. Cells which have been stored in this way are more spherical than normal, contain more sodium and potassium and their membranes are deficient in lipid. When injected into the circulation of a compatible recipient they are rapidly removed, mainly by the spleen. These cells are thus non-metabolizing and non-viable. If, however, the cells are taken and incubated for half to one hour with a suitable nucleoside a remarkable change occurs in them. They resynthesize adenosine triphosphate, and 2,3-diphosphoglycerate,

and begin again to metabolize glucose, so that if they are now injected into circulation their survival is greatly improved. These two changes are shown in Table 1. The mechanism through which this revival is brought about is probably the following.

TABLE 1. Percentage survival + ATP contents of cells stored for varying periods and incubated with nucleosides immediately before transfusion

Days of storage	Nucleoside	% survival at 48 hrs.		ATP content of cells ugP/ml. cells		
		Nucleoside	Glucose	Before incubation	After incubation Nucleoside	After incubation Glucose
28	Guanosine	75	54	25	54	30
	Adenosine	80	28	21	41	10
42	Guanosine	69	28	21	36	
	Adenosine	51	10	16	27	0
56	Guanosine	49	10	16	29	

When a cell's ATP stores have fallen below a critical level there is no longer sufficient present to assist the initial phosphorylation of glucose by hexokinase and the cell is apparently dead. If, however, the other source of sugar, the nucleoside, is present, ribose is split from the purine by a nucleoside phosphorylase and converted to ribose 5-phosphate which can then be utilized in the monophosphate shunt with the formation of triose phosphate and thence pyruvate, and the resynthesis of ATP; the reactions are made clear in Figure 1. In this way ATP is built up again in the cell and can once more be utilized for the phosphorylation of glucose. The apparently dead cell is now viable. The conclusion that can be drawn from these results is that the survival of the red cell depends on its being able to utilize glucose, and that this is dependent on sufficient cell stores of ATP.

These conclusions can be justified in another way, in this case by the use of metabolic inhibitors. Thus if blood is incubated with sodium arsenate for two hours the arsenate radicle replaces that of phosphate in esters within the cell; compounds of this sort can no longer be utilized and glycolysis stops. Cells treated in this way and labelled with an isotopic tag can be injected into the circulation, from which they rapidly disappear after

this treatment. There is one striking difference, however, between the destruction of cells in these two experiments; in the case of stored cells we have found that the cells are mainly destroyed in the recipient's spleen, whilst in the case of arsenolysed cells the destruction is throughout the circulation (Harris,

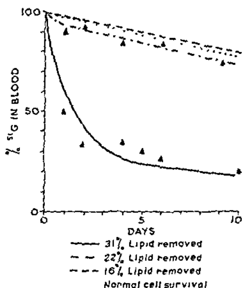


FIG. 2. Survival of ^{51}Cr-tagged red cells after various degrees of lipid depletion.

Prankerd and McAlister, 1957). It will be explained later that the principal factors in deciding the different sites of destruction of two such lots of cells are their thickness, the severity of their metabolic damage and their degree of agglutinability.

Another factor to be considered is the influence of the cell's surface on its survival. It has been mentioned that there are grounds for believing that the cell surface is not in a static condition but a dynamic one, this flux affects only the lipids, the protein component being stable. The problem can be approached by following the survival of cells after suitable treatment to remove a fraction of their lipid. In Figure 2 the cells involved were shaken with alumina for ten minutes during which time 20 per cent of their cholesterol and 15 per cent of

their phospholipid were removed, the total loss of lipid being about 31 per cent. On reinjection after labelling, the half-time of their survival was only 24 hours. Careful consideration of the cell survival curve shows that a fraction of the cells, about 60 per cent, was removed from the circulation rapidly whilst the remainder survived almost normally. This finding could be due to one of two things, either the cells present did not all lose lipid to the same extent, or else the lipid loss may have been uniform but those that survived normally were able to resynthesize their lipids after injection. The latter explanation is probably correct since after incubating the cells for an hour in their plasma before injection and after treatment with alumina, the number destroyed during the rapid phase is reduced. From a number of experiments of this nature it seems that about 20 per cent of cell lipid has to be removed in order to reduce their survival, less than this has negligible effect. It is difficult to visualize in what way the reduction in lipid leads to premature cell destruction. It might be that this treatment renders them mechanically more fragile, and this appears to be so *in vitro*, though what relationship shaking cells with glass beads has to their mechanical fragility in the circulation is problematical. Alternatively it might be that the removal of lipid alters the agglutinable properties of the cells and leads to their destruction by agglutination in situations where they stagnate, as has been so elegantly demonstrated by Jandl, Jones and Castle (1957) and Mollison (1959).

The next factor which must be considered is glutathione. The role of this compound in clinical haemolytic conditions will be discussed later, at this point only the experimental evidence that this substance can alter cell survival will be discussed. Fegler (1952) studied the haemolysis of horse erythrocytes *in vitro* on incubation after oxidation of their reduced glutathione with iodine. He was able to show a linear relationship between the amount of oxidized glutathione in the cell and in its liability to haemolysis under these conditions.

We have been particularly interested in the mechanism through which glutathione influences cell survival and Dr. Bailey has done some experiments which shed light on this problem.

Firstly blood was taken from a patient with favism whom it was known could not maintain the glutathione of his cells in a reduced state when they were incubated with acetyl-phenylhydrazine for 3 hours in the presence of oxygen; in this way cells containing oxidized glutathione were prepared. At the same time the same cells were incubated with acetyl-phenylhydrazine, but with the oxygen replaced by coal gas; and in these cells the glutathione remained in the reduced state. Cells were thus obtained containing oxidized or reduced glutathione, which had been treated similarly except for the presence of carbon monoxide which it is known does not alter the lifespan of red cells. Both types of cells were labelled with ^{51}Cr and injected into a compatible donor; their survival after injection is shown in Figure 3 where one can see the rapid disappearance of those containing oxidized glutathione. Various analyses were also made on the contents of these cells and the only change from normal found was a loss of lipid amounting to about 25 per cent in those cells containing oxidized glutathione. The effect of removal of lipid from red cells upon their lifespan has already been mentioned and it is possible that this is the mechanism through which a decrease in reduced glutathione operates to produce haemolysis.

Before considering the clinical problems, the processes accompanying red cell ageing must be discussed. Since the early studies of red cell survival curves it has been clear that the form of these is such as to indicate that each cell disappears after a lifespan of about 120 days. What change occurs in cells at this age which leads to their disappearance? The difficulty in tackling this problem is that analyses of samples of ordinary blood always consist of a mixture of cells of all ages, when what is required for analysis is a sample containing only very old or very young cells. Apart from age, however, the red cell population is also heterogeneous in its densities and in its osmotic fragilities and these two facts have been used to obtain more homogeneous samples of cells. If blood is obtained in which the young cells (i.e. a week old or younger) are labelled with ^{59}Fe, and this can readily be done by injecting the donor with ^{59}Fe and collecting the blood five days later, it can be shown that the youngest

cells are those which are less dense and least osmotically fragile (Prankerd, 1958; Marks, Johnson and Hirschberg, 1958). Because of this fact young cells can be separated at the top of a centrifuged column of cells, or as the residua after partial

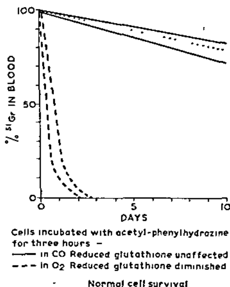


FIG. 3 Survival of ^{51}Cr-tagged red cells after oxidation of their glutathione.

haemolysis of a blood sample with hypotonic saline. Certain differences in cells segregated by these methods have now been shown. Marks and his colleagues using osmotic methods have shown a deterioration in enzyme activity as the cells age. It has also been shown, using centrifugal separation, that old cells contain less lipid than young cells (Prankerd, 1958; Westerman, Pierce and Jensen, 1959), but that there is no difference in their phosphate ester content. The two methods of separation agree in that the young cells from the top of a column of centrifuged cells are also most resistant to haemolysis in hypotonic saline.

Firstly blood was taken from a patient with favism whom it was known could not maintain the glutathione of his cells in a reduced state when they were incubated with acetyl-phenylhydrazine for 3 hours in the presence of oxygen; in this way cells containing oxidized glutathione were prepared. At the same time the same cells were incubated with acetyl-phenylhydrazine, but with the oxygen replaced by coal gas; and in these cells the glutathione remained in the reduced state. Cells were thus obtained containing oxidized or reduced glutathione, which had been treated similarly except for the presence of carbon monoxide which it is known does not alter the lifespan of red cells. Both types of cells were labelled with ^{51}Cr and injected into a compatible donor; their survival after injection is shown in Figure 3 where one can see the rapid disappearance of those containing oxidized glutathione. Various analyses were also made on the contents of these cells and the only change from normal found was a loss of lipid amounting to about 25 per cent in those cells containing oxidized glutathione. The effect of removal of lipid from red cells upon their lifespan has already been mentioned and it is possible that this is the mechanism through which a decrease in reduced glutathione operates to produce haemolysis.

Before considering the clinical problems, the processes accompanying red cell ageing must be discussed. Since the early studies of red cell survival curves it has been clear that the form of th [illegible] lifes[illegible] age [illegible] ling this problem is that analyses of samples of ordinary blood always consist of a mixture of cells of all ages, when what is required for analysis is a sample containing only very old or very young cells. Apart from age, however, the red cell population is also heterogeneous in its densities and in its osmotic fragilities and these two facts have been used to obtain more homogeneous samples of cells. If blood is obtained in which the young cells (i.e. a week old or younger) are labelled with ^{59}Fe, and this can readily be done by injecting the donor with ^{59}Fe and collecting the blood five days later, it can be shown that the youngest

cells are those which are less dense and least osmotically fragile (Prankerd, 1958; Marks, Johnson and Hirschberg, 1958). Because of this fact young cells can be separated at the top of a centrifuged column of cells, or as the residue after partial

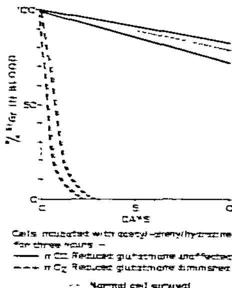


Fig. 3. Survival of ^{51}Cr-tagged red cells after oxidation of their glutathione.

haemolysis of a blood sample with hypotonic saline. Certain differences in cells segregated by these methods have now been shown. Marks and his colleagues using osmotic methods have shown a deterioration in enzyme activity as the cells age. It has also been shown, using centrifugal separation, that old cells contain less lipid than young cells (Prankerd, 1958; Westerman, Pierce and Jensen, 1959), but that there is no difference in their phosphate ester content. The two methods of separation agree in that the young cells from the top of a column of centrifuged cells are also most resistant to haemolysis in hypotonic saline. This is all that appears to be known at present of changes occurring in ageing red cells: it is a problem which presents an important challenge to research.

Firstly blood was taken from a patient with favism whom it was known could not maintain the glutathione of his cells in a reduced state when they were incubated with acetyl-phenylhydrazine for 3 hours in the presence of oxygen; in this way cells containing oxidized glutathione were prepared. At the same time the same cells were incubated with acetyl-phenylhydrazine, but with the oxygen replaced by coal gas; and in these cells the glutathione remained in the reduced state. Cells were thus obtained containing oxidized or reduced glutathione, which had been treated similarly except for the presence of carbon monoxide which it is known does not alter the lifespan of red cells. Both types of cells were labelled with ^{51}Cr and injected into a compatible donor; their survival after injection is shown in Figure 3 where one can see the rapid disappearance of those containing oxidized glutathione. Various analyses were also made on the contents of these cells and the only change from normal found was a loss of lipid amounting to about 25 per cent in those cells containing oxidized glutathione. The effect of removal of lipid from red cells upon their lifespan has already been mentioned and it is possible that this is the mechanism through which a decrease in reduced glutathione operates to produce haemolysis.

Before considering the clinical problems, the processes accompanying red cell ageing must be discussed. Since the early studies of red cell survival curves it has been clear that the form of these is such as to indicate that each cell disappears after a lifespan of about 120 days. What change occurs in cells at this age which leads to their disappearance? The difficulty in tackling this problem is that analyses of samples of ordinary blood always consist of a mixture of cells of all ages, when what is required for analysis is a sample containing only very old or very young cells. Apart from age, however, the red cell population is also heterogeneous in its densities and in its osmotic fragilities and these two facts have been used to obtain more homogeneous samples of cells. If blood is obtained in which the young cells (i.e. a week old or younger) are labelled with ^{59}Fe, and this can readily be done by injecting the donor with ^{59}Fe and collecting the blood five days later, it can be shown that the youngest

the use of substrates other than glucose such as nucleosides. This would suggest that the defect, if due to an enzyme involved in glycolysis, is located between the phosphorylation of glucose and the formation of triose phosphates. The hereditary spherocytic cell is also unusually sensitive to the inhibitory action on metabolism of fluoride, and an enzyme that would satisfy these criteria would be phosphofructokinase. The location here or elsewhere awaits further proof. A question that it is important to consider is the importance of the glycolytic defect in the natural history of this disease. It is well known that the haemolytic component of the disease can be removed by splenectomy with apparent cure, but without alteration of the cell abnormality. This might suggest that it is the shape of the cells which is the determining factor in their destruction since there is abundant evidence now that spherocytic cells are sequestered in the spleen. An explanation in terms of cell shape, however, neglects two phenomena which are important for a fuller understanding of the disease process. Firstly, it is possible to produce spherocytic cells by suitably heating blood at about 50° C. for a few minutes, and this degree of heating does not appear to cause any metabolic damage although after this treatment the cells are permanent spherocytes. By altering the conditions of heating it is possible to produce blood with fragility curves mimicking those of the natural disease, but in this case with normal metabolic functions (Fig. 4). If these cells are labelled and injected into a compatible recipient it is interesting to compare the patterns of their survival with those of the natural spherocytes similarly labelled. Figure 5 shows the differences that are found. Splenic equilibration curves are similar with both types of cell, but the daily accumulation of radioactivity in the spleen is less in the case of heated cells, whilst their elimination from the peripheral blood is also less. Since the only detectable differences between the cells appear to be in their metabolism it is reasonable to suppose that it is this metabolic difference that accounts for the different patterns of survival.

A feature of the disease which also demonstrates the importance of the metabolic defect is the occurrence of a form in which the osmotic fragility of fresh cells is normal. These cells therefore

The dynamic chemical and physical factors considered so far have omitted any reference to haemoglobin. This is partly because haemoglobin does not take any active part in cell metabolism, except through such mechanisms as keep it in reduced state, and partly because nothing is known of the way in which a cell reaches its premature fate when possessing an abnormal haemoglobin, except in the case of the sickle pigment. In the latter case the gelling of the haemoglobin within the cell both distorts the shape and stops metabolism and in this state the cell cannot survive long.

CHEMICAL AND PHYSICAL FACTORS IN CLINICAL HAEMOLYSIS

The first group of diseases to consider are the congenital haemolytic anaemias, and the common example of this is hereditary spherocytosis, or as it is still sometimes called, acholuric jaundice. There are two defects in the cells of this disease. The first, their shape is well known, although not all cases have spherocytosis in fresh blood. The cause of the spherocytosis is not known, but it is probably related to the other cellular defect, that involving glycolysis. The link between the two abnormalities could be that the enzyme involved in metabolism is also a structural component of the cell, or, if the cell shape is the result of a dynamic state involving a contractile protein, failure of contracture could result from the glycolytic defect. So far it is only possible to define the metabolic defect from the relationship of ^{32}P fluxes in the various phosphate esters in the cell, but this

TABLE 2 Mean relative specific activities of phosphate esters of red cells after incubation with ^{32}P for 4 hours

	ATP	Inorganic P
Normal cells	336 ± 6 8	384 ± 10 4
H.S cells	221 ± 16 4	460 ± 30
Diff of means ± SE	115 ± 18	112 ± 32
	$t < 0\ 001$	$t < 0\ 01$

abnormality is very marked and is illustrated in Table 2. These findings have been confirmed by other workers, and it has been found that the defective phosphorylation can be corrected by

the use of substrates other than glucose such as nucleosides. This would suggest that the defect, if due to an enzyme involved in glycolysis, is located between the phosphorylation of glucose and the formation of triose phosphates. The hereditary spherocytic cell is also unusually sensitive to the inhibitory action on metabolism of fluoride, and an enzyme that would satisfy these criteria would be phosphofructokinase. The location here or elsewhere awaits further proof. A question that it is important to consider is the importance of the glycolytic defect in the natural history of this disease. It is well known that the haemolytic component of the disease can be removed by splenectomy with apparent cure, but without alteration of the cell abnormality. This might suggest that it is the shape of the cells which is the determining factor in their destruction since there is abundant evidence now that spherocytic cells are sequestered in the spleen. An explanation in terms of cell shape, however, neglects two phenomena which are important for a fuller understanding of the disease process. Firstly, it is possible to produce spherocytic cells by suitably heating blood at about 50° C. for a few minutes, and this degree of heating does not appear to cause any metabolic damage although after this treatment the cells are permanent spherocytes. By altering the conditions of heating it is possible to produce blood with fragility curves mimicking those of the natural disease, but in this case with normal metabolic functions (Fig. 4). If these cells are labelled and injected into a compatible recipient it is interesting to compare the patterns of their survival with those of the natural spherocytes similarly labelled. Figure 5 shows the differences that are found. Splenic equilibration curves are similar with both types of cell, but the daily accumulation of radioactivity in the spleen is less in the case of heated cells, whilst their elimination from the peripheral blood is also less. Since the only detectable differences between the cells appear to be in their metabolism it is reasonable to suppose that it is this metabolic difference that accounts for the different patterns of survival.

A feature of the disease which also demonstrates the importance of the metabolic defect is the occurrence of a form in which the osmotic fragility of fresh cells is normal. These cells therefore

cannot be said to be spherocytic. On incubation *in vitro*, however, they behave like the cells of hereditary spherocytosis in that they swell significantly more than normal cells so that their osmotic fragility after 24 hours' incubation is increased. A case of

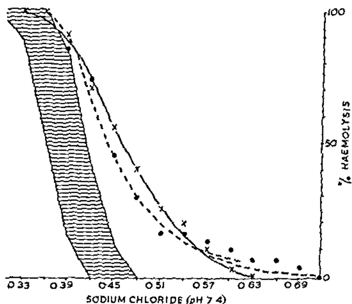


FIG. 4. Osmotic fragilities of heated cells and natural spherocytes.
x——x Hereditary spherocytic cells
● - - - ● Heated cells

this type has been studied; the cells showed the typical metabolic defect and the patient had a moderately severe haemolytic state with a serum bilirubin of 1·7 mg. per cent, 14 per cent reticulocytes and a mean cell survival of about 20 days. The red cells were labelled and injected into a compatible normal recipient and as there was no spherocytosis there was no slow equilibration curve of radioactivity over the spleen as seen when spherocytes are sequestered there; however, over the following days the radioactivity in the spleen became greater than normal and continued to rise, and at the same time the cells were removed rapidly from the circulation (Fig. 6). How can we explain the splenic destruction in this instance? There is plenty of evidence

to show that normal cells stagnate several hours in the spleen and that during this time they swell and gain water and electrolytes. The swelling of cells in the spleen from this particular case would be greater than normal if the *in vitro* results are analogous

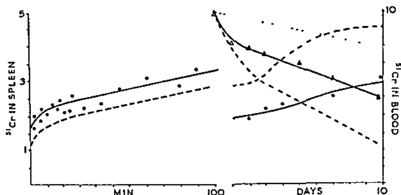


FIG 5 Survival of ^{51}Cr-tagged red cells and splenic radioactivity after injection of heated and H.S red cells of similar osmotic fragility (two normal recipients)

▲ ▲ Blood counts { ——— heated cells; - - - - - H S. cells }

● ● Spleen counts { ——— heated cells; - - - - - H.S. cells }

to conditions in the spleen and this excessive swelling would then result in the cells becoming trapped in the spleen. There is one further question we must ask and that is why the excessive swelling of the hereditary spherocytic cells occurs. If their metabolism is defective one might expect their ability to maintain the normal electrolyte gradients to be defective also, and it has recently been found that on incubating spherocytes without glucose for 3 hours they undergo greater swelling and show a more rapid breakdown of their phosphate esters than do normal cells (Prankerd, 1960). In particular 2,3-diphosphoglycerate breaks down more quickly.

packed together and the amount of available substrate reduced;

as a result of this swelling their transit through the spleen is further delayed, packing accentuated, and a vicious circle brought into operation. Furthermore, cells which escape from the spleen after a period of swelling may after recirculation re-enter the spleen and undergo further stagnation and chemical change.

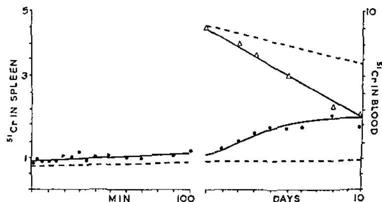


Fig 6 Survival of ^{51}Cr-tagged red cells and splenic radioactivity after injection of non-spherocytic H S cells into normal recipients.

△——△ Blood counts (H.S. cells)
●——● Spleen counts (H.S. cells)
- - - - - - Normal red cells

Apart from hereditary spherocytosis there are a number of other congenital haemolytic anaemias. In some the cell morphology is abnormal, and in some biochemical defects in the cells can be demonstrated. As yet this remains an ill-defined group of disorders, but we have recently examined the cells from a hereditary haemolytic anaemia in a young boy and his father, neither of whom was cured by splenectomy and who both showed a heterogeneous population of cells, some thicker and some thinner than normal. In these we have been able to show a defective synthesis of 2,3-diphosphoglycerate presumably due to a deficiency of the enzyme phosphoglyceric acid mutase. In other anaemias of this group enzyme defects probably also exist. It may seem odd that some cells with a certain biochemical lesion are destroyed in the spleen whilst others with a different

lesion are destroyed throughout the body. One might ask what it is about the chemical defect that determines the site of destruction. The exact answer to this question is unknown; undoubtedly cell thickness plays a part and if the cells are abnormally thin, then when under metabolic stress, they can undergo more swelling before being trapped in the spleen. Another factor which is probably important is the severity of the metabolic defect; it has already been indicated how non-metabolizing cells are destroyed throughout the circulation and it seems reasonable that when the defect is severe cell destruction will be more widespread, and of course splenectomy less effective as treatment.

Haemolytic anaemias associated with a deficiency of the enzyme glucose 6-phosphate dehydrogenase have been extensively investigated recently. These anaemias were discovered as a result of the haemolysis observed after prophylactic treatment of malaria in negroes with Primaquine. A group of workers in Chicago (Beutler, Dern, Flanagan and Alving, 1955) have investigated this phenomenon and shown that as a result of the enzyme deficiency these individuals cannot maintain their red cell glutathione effectively in a reduced state; in the presence of Primaquine glutathione becomes irreversibly oxidized. The haemolytic state produced by Primaquine is self-limiting and these workers have also shown that this is due to a difference in the enzyme defect between young and old cells; once the older cells with the greater defect have been eliminated the rate of haemolysis declines. Patients are thus no longer susceptible to haemolysis immediately after an attack if they are given more of the drug. This is very elegantly shown in Figure 7. Beutler (1957) has devised a test for glutathione stability which proves a simple means of detecting subjects with susceptible red cells. In this test blood is incubated with acetyl phenylhydrazine and the decrease in reduced glutathione estimated.

Sheba and his colleagues in Israel (Szeinberg, Sheba, Hirshorn and Bodonyi, 1957) have now shown that the same defect occurs in many Asiatic Jews and is responsible for the occurrence of haemolysis in favism, and after taking such drugs as sulphonamides and P.A.S. Other workers have shown that

as a result of this swelling their transit through the spleen is further delayed, packing accentuated, and a vicious circle brought into operation. Furthermore, cells which escape from the spleen after a period of swelling may after recirculation re-enter the spleen and undergo further stagnation and chemical change.

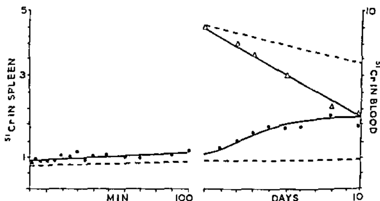


FIG 6. Survival of ^{51}Cr-tagged red cells and splenic radioactivity after injection of non-spherocytic H S. cells into normal recipients.

△——△ Blood counts (H.S cells)
●——● Spleen counts (H S cells)
- - - - - - Normal red cells

Apart from hereditary spherocytosis there are a number of other congenital haemolytic anaemias. In some the cell morphology is abnormal, and in some biochemical defects in the cells can be demonstrated. As yet this remains an ill-defined group of disorders, but we have recently examined the cells from a hereditary haemolytic anaemia in a young boy and his father, neither of whom was cured by splenectomy and who both showed a heterogeneous population of cells, some thicker and some thinner than normal. In these we have been able to show a defective synthesis of 2,3-diphosphoglycerate presumably due to a deficiency of the enzyme phosphoglyceric acid mutase. In other anaemias of this group enzyme defects probably also exist. It may seem odd that some cells with a certain biochemical lesion are destroyed in the spleen whilst others with a different

6-phosphate dehydrogenase have an excess of glutathione reductase, a finding which suggests an attempt to compensate for the diminished efficiency of glutathione reduction.

The last factor to be considered is the role of cell lipids in haemolytic states. Here the ground is at present still insecure. Erickson and her colleagues in 1937 analysed the lipid content of red cells from a number of patients with hereditary haemolytic disorders and found rather variable results. They expressed their results as quantities of lipid per cell, but when one considers the functional significance of lipid in the cell it is really the concentration per unit area which is important. If Erickson's data are recalculated in this form certain differences between types of cells are clearer. We have repeated this type of analysis and found somewhat similar results to Erickson's; a striking abnormality amongst various cells studied being a deficiency of cerebroside in the thalassaemic cell.

Another haemolytic disease which looks as if it may be primarily an error of lipid synthesis is paroxysmal nocturnal haemoglobinuria. Harris, Prankerd and Westerman (1957) have found the cells in this condition deficient in lecithin yet with a compensatory increase in phosphatidyl serine, and Munn and Crosby (1957) have reported certain abnormalities in the fatty acid constitution of PNH cells. Recently Lovelock and I have been able to find similar differences on separating the fatty acids by gas chromatography. However, Barry (1959) has found no change in cell phospholipids, whilst Hartmann and Anditore (1959) have reported a deficiency of acetyl choline esterase in these cells. These conflicting reports remain to be clarified. It has long been known that the primary defect in this disease resides in the red cells rendering them more susceptible to naturally existing haemolytic systems in the plasma, which in this case includes complement, properdin and magnesium ions.

This completes the present survey of the life of the red cell. An attempt has been made to indicate that the cell is a living component of the body, and that in order to understand the mechanisms responsible for its destruction, we must first discover the means whereby it maintains its peculiar existence in

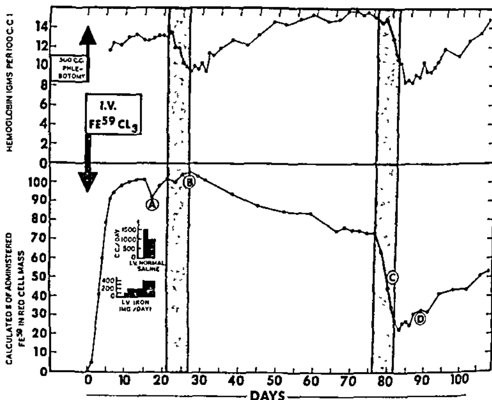


FIG 7. The effect of primaquine administration upon an Fe^{59}-labelled red cell population with a narrow age range (Shaded areas represent 6-day courses of primaquine) (From *J Lab Clin Med*, 44, p 440, Fig 1)

the same defect is also responsible for the occurrence of haemolysis in certain people given nitrofurantoin, children eating mothballs and those infants who sometimes show haemolysis after synthetic vitamin K injections. It is obvious that the glutathione mechanism is becoming of widespread importance in haemolysis complicating the administration of drugs. All workers have concluded that the enzyme defect is genetically determined and Gross, Hurwitz and Marks (1958) have produced evidence to show that its transmission is sex-linked. Other recent work shows that the cells which are deficient in glucose

XVII

The Role of Biochemistry in Medicine

M. MAIZELS

Discussion of such a topic as the role of biochemistry in medicine is less simple than might be thought: it is as though one had to debate the place of history in education, though even here opinions differ, for no less an authority than the late Henry Ford has declared that history is bunk. A better parallel to our subject is the contribution of plumbing to civilization, and here an answer of a sort may be reached by imagining the modern scene in the absence of modern plumbing. So too an answer to our present problem may be reached, by considering what medicine was like at the dawn of the biochemical era. This in turn involves a consideration of the contributions of biochemistry to clinical medicine.

In the millennia that preceded the eighteenth century, though many contributions were made to the knowledge of the structure of the body, little was understood of function, and it is probably true to say that the first major contribution to physiology was Harvey's discovery of the circulation of the blood in 1657. It is not surprising, therefore, that the treatment of disease, lacking any scientific foundation, was determined by the personal opinions and prejudices of the physician, and was, to some extent, dictated by the general belief in magical spells. An example is Sir Robert Boyle's remedy for colic (1725): 'Take four or five balls of fresh horse dung, and let them steep for about a quarter of an hour in a pint of white wine, . . . strain and give of it from a quarter to half a pint, or more at a time; the patient having a care not to take cold after it.' By contrast,

the circulation; these means are mainly chemical and physical, and it is in these fields that a closer understanding of haemolytic mechanisms probably resides.

ACKNOWLEDGEMENT

Figs. 4, 5 and 6 from *Quart. J. Med.* (1960), **29**, 155 with permission of the publishers.

REFERENCES

ALTMAN, K. I. (1951). *Arch. Biochem* 33, 169.
BARRY, R. M. (1959). *Brit. J. Haematol.* 5, 212.
BEUTLER, E. (1957). *J. Lab. Clin. Med.* 48, 84.
BEUTLER, E., DERN, R. J., FLANAGAN, C. L. and ALVING, A. S. (1955). *J. Lab. Clin. Med.* 45, 286.
BUCHANAN, A. (1960). *Biochem J* 75, 315.
ERICKSON, B. N, WILLIAMS, H. H., HUMMEL, F. C., LEE, P. and MACEY, I. G (1937). *J. biol. Chem.* 118, 569.
FEGLER, G. (1952). *Nature, Lond.* 170, 624.
GROSS, R. T., HURWITZ, R. E and MARKS, P. A (1958). *J. clin. Invest.* 37, 1176
HARRIS, I. M., PRANKERD, T. A. J. and MCALISTER, J. (1957). *Clin. Sci.* 16, 223.
HARRIS, I. M., PRANKERD, T. A. J. and WESTERMAN, M. P. (1957). *Brit med. J.* ii, 1276.
HARTMANN, R. and ANDITORE, J. V (1959). *J. clin Invest.* 38, 1009.
JAMES, A T., LOVELOCK, J E. and WEBB, J. (1957). *Proc. Biochem. Soc.* 66, 60P.
JANDL, J. H., JONES, A. R. and CASTLE, W B (1957). *J. clin. Invest* 36, 1428.
MARKS, P. A., JOHNSON, A. B. and HIRSCHBERG, E. (1958). *Proc. Nat. Acad. Sci.* 44, 529
MOLLISON, P. L. (1959). Oliver Sharpey Lectures, Royal College of Physicians
MUNN, J. I. and CROSBY, W H (1957). *Proc Soc. exper. Biol Med.* 96, 480.
PRANKERD, T. A. J. (1956) *Internat Rev Cytol.* v, 279
PRANKERD, T. A. J. (1958). *J Physiol.* 143, 325.
PRANKERD, T. A. J. (1960). *Quart. J Med* 29, 155
SZEINBERG, A., SHEBA, C, HIRSHORN, N. and BODONYI, E. (1957) *Blood,* 12, 603.
WESTERMAN, M. P., PIERCE, L. E. and JENSEN, W. N. (1959). *J clin. Invest.* 38, 1054.

XVII

The Role of Biochemistry in Medicine

M. MAIZELS

Discussion of such a topic as the role of biochemistry in medicine is less simple than might be thought: it is as though one had to debate the place of history in education, though even here opinions differ, for no less an authority than the late Henry Ford has declared that history is bunk. A better parallel to our subject is the contribution of plumbing to civilization, and here an answer of a sort may be reached by imagining the modern scene in the absence of modern plumbing. So too an answer to our present problem may be reached, by considering what medicine was like at the dawn of the biochemical era. This in turn involves a consideration of the contributions of biochemistry to clinical medicine.

In the millennia that preceded the eighteenth century, though many contributions were made to the knowledge of the structure of the body, little was understood of function, and it is probably true to say that the first major contribution to physiology was Harvey's discovery of the circulation of the blood in 1657. It is not surprising, therefore, that the treatment of disease, lacking any scientific foundation, was determined by the personal opinions and prejudices of the physician, and was, to some extent, dictated by the general belief in magical spells. An example is Sir Robert Boyle's remedy for colic (1725): 'Take four or five balls of fresh horse dung, and let them steep for about a quarter of an hour in a pint of white wine, . . . strain and give of it from a quarter to half a pint, or more at a time; the patient having a care not to take cold after it.' By contrast,

the nineteenth century was a period of great scientific advance in anatomy, histology, physiology and bacteriology, though biochemistry was still unknown. At that time, the nature of many disease processes seemed clear, but the new knowledge could only be translated into action in the case of a few diseases: arsenicals for syphilis, iron for anaemia, antitoxin for diphtheria and so on. In the vast majority of illnesses, rest in bed and good nursing were all that could be offered, other than symptomatic treatment. Since it must be conceded that the ultimate aim of medicine is to cure patients, the situation was frustrating, and in an atmosphere of much apparent knowledge and few practical applications, it is not surprising that the magical element in medicine, inherited from previous centuries, survived and was for a while even intensified, so that as a short-cut to success in the early part of the present century it was necessary for a doctor to 'shoot a line'. In some cases, this line had no factual basis at all: in others—and these the more dangerous—it was based on existing knowledge, but was carried far beyond what was justified by the evidence, the criterion being that because an idea was reasonable, or at least plausible, therefore it was probably true. At this point two matters must be emphasized: first, the 'line-shooting' witch-doctor genuinely believed in his own theories (as in fact all the best witch-doctors do); he did not consciously cheat, but deceived himself as much as he deceived his patients. Second, the great bulk of the medical profession, consultant and general practitioner alike, was content even at that time to serve honestly and sincerely within the then defined limits of medical knowledge; if success were attained, it was achieved through qualities of mind and character. However, the first group, though small, was disproportionately influential because it was more vocal, more histrionic, and its members were in a word the showmen of the profession.

It may be wondered what forms were taken by medical speculation. They were many and various. Thus forty years ago, the lithaemic diathesis was still in vogue and included a great variety of conditions all supposed to have some common gouty factor: such conditions were chronic eczema, arteriosclerosis, cirrhosis, and of course, gout itself. What the diagnosis of the

lithaemic diathesis really implied remains obscure. The tuberculous diathesis and the tuberculides were other vague entities, and of course there was allergy, which still remains a fruitful field for exploitation. Introduced a little later, and based on the new bacteriological knowledge, was the concept of focal sepsis, a popular focus being the large bowel. Toxic absorption from this site was thought to occur in rheumatoid arthritis, exophthalmic goitre, chronic eczema, arteriosclerosis and many other chronic conditions. The condition might be treated by a prolonged course of sour milk, or by a course of autogenous vaccine prepared from the patient's intestinal flora, or by colectomy: procedures to which the medical attendant was the more enjoined if he could show that there was an excess of indican in the urine; for indicanuria was regarded as an index of intestinal putrefaction. There was of course no sort of evidence that either constipation or excess of indican in the urine is associated with the maladies in question, or indeed is really harmful at all: but all this was beside the point; the idea was plausible—therefore it was true. Looking back after the lapse of many years, it seems remarkable that drastic procedures like colectomy for rheumatoid arthritis and goitre could be carried out, without any attempt at a statistical analysis of the results obtained.

With the pioneer work of Henderson and of Van Slyke in the second decade of this century, acidosis and alkalosis became familiar terms, though their precise implications were not always well understood. In spite of this, the terms acid and alkaline diathesis became commonplace, and were used in default of any supporting evidence, which was neither wanted nor sought. From this period date such catch-phrases as the following. 'Youth is sthenic, tending to acidosis and dental caries: age is asthenic, tending to alkalosis and pyorrhoea', 'Seborrhoeic dermatitis is due to acidosis, the acid concerned being hippuric acid' (100 per cent of a series of one case). So too in a reputable textbook on pediatrics, published as recently as 1938, there appears a diagram (vaguely reminiscent of Clapham Junction) which shows that rickets, tetany, infantile eczema, endocrine disorders, acute and chronic infections, cyclical vomiting and other disorders, are either the causes or the results of acidosis in

the young. It is perhaps unnecessary to add that all these statements are wholely or very largely untrue.

That the preceding statements are not the outcome of subjective exaggeration may be shown by a consideration of the impact of early twentieth-century medicine on the intellectuals of the day. Two quotations from the mordant preface to Bernard Shaw's *The Doctor's Dilemma* will suffice: '. . . the medical service of the community, as at present provided for, is a murderous absurdity', and '. . . he draws disastrous conclusions from his clinical experience because he has no conception of scientific method . . .'.

It may well be wondered why it is that today witchcraft and magic have largely disappeared from medicine. The reason is twofold: first, many more diseases can now be adequately treated: diabetes with insulin, pernicious anaemia with vitamin B_{12}, infections with antibiotics, the so-called collagen diseases with cortisone, and so on. The second factor is the amazing advance of biochemistry. Biochemistry has exerted its influence in several ways: it has provided methods of investigation, and these methods have been the machine-tools of research. Research in its turn has provided an explanation of the chemical basis of disease on the one hand, and on the other has suggested further methods and tests which are today the foundations of diagnosis, prognosis and treatment. All these are positive achievements; and as a negative contribution it has 'debunked' the facile and specious speculations of an earlier generation. This negative contribution should not be underestimated.

Mention has already been made of the importance of methods in biochemistry: for clinical work these methods must be reasonably quick, not too exacting, and sparing of pathological material. Hence it is, that in this field, the micro-methods of Folin, Benedict, Van Slyke and King are of special value, and should rank in importance with their other work. We are too apt to take our methods for granted, so that Van Slyke is not now a man but a manometer, while Folin is merely the parttitle of a method already osbolete and almost forgotten.

Biochemistry also adapts the techniques of chemistry and physics, and these methods are further adapted to the needs of

the medical laboratory. Recent acquisitions of fundamental importance are flame-photometry, paper electrophoresis and chromatography, and the use of radioactive isotopes. Hence it is that in 1920 blood chemistry was restricted to ureas and sugars, that even in 1940 estimations of the alkali reserve, of plasma phosphorus, calcium, sodium, potassium and proteins were regarded as tiresome necessities, supportable only because they were so rarely requested, while today all these are regarded as commonplace and essential tests.

Referring yet again to the importance of methods one recalls the work of Zuelzer, who between 1905 and 1908 investigated the effects of injecting pancreatic extracts into animals. He showed quite clearly (1908) that such injections into pancreatectomized dogs greatly decreased ketosis and glycosuria, though the ultimate mortality was high, some of the animals dying in fits. It is probable that most of the deaths were brought about by the crude extracts used, but some may well have been due to hypoglycaemia, a contingency which Zuelzer could hardly have envisaged in the absence of accurate micromethods for estimating blood sugars; a decade or so later investigation of the problem using extracts not of whole pancreas but of islet tissue, and combining this with appropriate biochemical techniques, led Banting and Best to the epoch-making discovery of insulin. So, too, the introduction of flame-photometry with its rapid and simple estimation of sodium and potassium has vastly extended our knowledge of the factors concerned in salt balance, and greatly facilitated the investigation and treatment of conditions where the balance of sodium and potassium has been disturbed. It may be well at this point to consider some examples of the value of these and other tests in the treatment of disease. These have been chosen for their interest and there is no more connexion between one item and another than there is between the various poems in an anthology.

Potassium Depletion and Excess

Forty years ago diabetic coma was almost invariably fatal. When the association of coma with ketosis was appreciated, alkaline therapy was instituted, but the survival rate was not

improved. Indeed, even after the introduction of insulin many cases died unexpectedly in spite of the relief and control of acidosis and hyperglycaemia. It was then shown that prolonged diuresis and acidosis caused depletion of salts and water, and predisposed to a shock-like condition with low blood pressure. It will be recalled that in diabetes excess of acids in the urine is combined with ammonia formed in the kidneys. But in severe acidosis this mechanism may become inadequate, and excessive acidity of the urine is then prevented by a call on the fixed bases of the body. Hence it is that salt and water depletion occur with the supervention of shock. Even so, treatment with copious infusions of sodium chloride solution was often unsuccessful, and it was ultimately realized that an important factor in cation depletion was not only loss of sodium, but also loss of potassium, and that death when it occurred was often due to hypokalaemia. It may be noted that in this, as in other forms of extrarenal acidosis, phosphate depletion coexists. It arises as follows: there is an equilibrium between phosphoric esters and inorganic phosphate within the cells depending on the balance of phosphorylation and dephosphorylation. If the cell becomes more alkaline, the balance shifts in favour of phosphorylation: if the cell becomes more acid, it shifts in favour of dephosphorylation; cell phosphoric esters decrease and cell inorganic phosphate tends to increase. There is, however, a further equilibrium between inorganic phosphate within the cells and that outside in the extracellular fluid. Hence when inorganic phosphate within the cells tends to rise, it leaks into the extracellular medium and thence into the blood stream, whence it is lost from the body by excretion in the urine. Changes in total cell phosphate with variations in the alkali reserve are shown in Table 1. It follows that in acidosis, phosphate capable of binding cell base is lost from the body. For a fall in pH of 0·1 units this may easily amount to 7 mM phosphate capable in the case of erythrocytes of binding 12 m.equiv cation per l. cells, further, with fall of pH haemoglobin or other cell protein binds less cation, and the total decrease in cation bound by red cells for 0·1 unit fall of pH may be about 17 m equiv/l. or about 14 per cent of the total cation present. For other body cells the findings will be

qualitatively similar, and there will be a loss of cation from the tissues in general. Since about 65 per cent of cell cation is potassium, acidosis tends towards depletion of tissue potassium. Hence in acidosis, though the necessity for excreting cation to combine with excess acids in the urine involves both sodium and potassium, the lability of tissue potassium makes it specially vulnerable. It should be noted that the successful treatment of the acidosis with alkaline fluids *devoid* of potassium will restore the complement of phosphoric esters in the cells and also the base-binding power of cell proteins, so that with the relief of acidosis, the effects of hypokalaemia may become even more prominent.

TABLE 1. Alkali reserve and phosphate in red cells

	Plasma CO_2 vols. %	Total cell phosphate m.equiv/l. cells
Acidosis	22	20
Recovery	65	29
Alkalosis	89	39
Recovery	58	31

Apart from specific effects of sodium and potassium deficiencies, there also occur in diabetic coma the general effects of salt loss. Thus excessive loss of sodium in the acid urine tends to lower the osmotic pressure of the intercellular fluids and a compromise compensation may lead to a concomitant loss of water, limiting the fall of osmotic pressure but leading to dehydration: so too, loss of potassium from the tissues must involve loss of tissue water and so lead to dehydration of the tissue cells. These findings are the basis of therapy by Butler's solution, which contains water and sodium chloride for the relief of dehydration, lactate whose oxidation provides bicarbonate for the relief of acidosis, and potassium and phosphate to make good phosphate and potassium depletion. Vigorous rational treatment on these lines, together with the administration of insulin and possibly of intravenous glucose, have transformed the prognosis in diabetic coma, but it does require careful control of the blood sugar by chemical methods, of the alkali reserve by means of the Van Slyke apparatus, and of plasma sodium and potassium by

flame-photometry. As a corollary, it follows that no institution is equipped to deal with this emergency unless a day and night biochemical service is available. These implications follow from the valuable work of Gamble, Hartmann and others in the United States during the past two decades, but it should not be forgotten that thirty-five years ago Haldane, Wigglesworth and Woodrow (1924), experimenting with the effects of ammonium chloride feeding in man, remarked on the acidosis with its concomitant loss of cell phosphate and potassium, and stated explicitly that in prolonged acidosis, phosphate and potassium depletion might have serious effects.

Many other conditions may be associated with hypokalaemia and all of them require that therapy be controlled with flame photometry. Among these are severe and prolonged vomiting or diarrhoea from any cause, and it is well to remember that continued gastric aspiration with or without lavage may lead not only to alkalosis and hypochloraemia, but also to hypokalaemia. So, too, in intestinal ileus, the coils of intestine may be greatly distended with fluid rich in salts, so that sodium and potassium deficiency may occur without any vomiting or diarrhoea at all. *Hypokalaemia may also occur from the prolonged exudation of* burns, in treatment with cortisone and ACTH, and it is sometimes induced along with acidosis when base-binding resins are given orally to restrict sodium intake in cases of oedema or high blood pressure. *On the other hand, hyperkalaemia* may be equally serious. It occurs in association with shock and secondary renal damage after crush injury, and it may be delayed by the application of a tourniquet (where this is feasible). Hyperkalaemia may also occur after copious blood transfusion with stored blood, and hence is not unknown in the replacement transfusion of erythroblastosis foetalis. It is frequent during the anuria of incompatible blood transfusion, but it may be succeeded by hypokalaemia if the patient recovers and copious diuresis supervenes. Among other causes of hyperkalaemia are deficiency of adrenal cortical hormones or of cortico-trophic hormone. There are other causes of potassium deficiency and excess besides those already mentioned, but enough has been said on this topic to show how flame-photometry has made us

all familiar with sodium and potassium disturbances and their results, and has enabled us to treat what otherwise would be serious and even fatal complications of many surgical and medical conditions. The following is an interesting clinical example: a man was admitted to hospital febrile and semi-conscious, with slight head-retraction. His condition suggested subarachnoid haemorrhage, and his cerebrospinal fluid was indeed somewhat blood-stained, though otherwise normal, except for the chloride which was 95 instead of 120 m.equiv/l. Blood was therefore tested and showed a normal alkali reserve with a marked chloride and cation deficit: plasma sodium was low at 112 m.equiv and potassium high at 7 8 m.equiv. A diagnosis of acute Simmonds disease was made, and so it turned out, for a large haemorrhage involving the pituitary was found post-mortem. Simmonds disease with such an acute onset is very rare: it usually occurs in females in relation to severe antepartum or postpartum haemorrhage, and tends to run a more chronic course.

Paper Chromatography and Paper Electrophoresis

By means of paper electrophoresis it has become possible to identify a host of different haemoglobins having an important bearing on anaemia, and also a number of serum proteins, which with electrophoresis on starch may number thirteen or more. Not infrequently a diagnosis of collagen disease or multiple myelomatosis is suggested by a high γ-globulin on simple electrophoresis of serum on paper, and this and associated laboratory tests are now rightly used in any case of obscure pyrexia.

Paper chromatography has enabled us to identify simply and quickly lipids in tissues and reducing substances in urine. It has many other applications, including the demonstration of amino-aciduria and hence of renal tubular defect in certain types of resistant rickets. These are discussed in the next section.

Resistant Rickets

This rare group differs from renal rickets associated with chronic nephritis in that the usual evidence of nephritis is lacking: the

flame-photometry. As a corollary, it follows that no institution is equipped to deal with this emergency unless a day and night biochemical service is available. These implications follow from the valuable work of Gamble, Hartmann and others in the United States during the past two decades, but it should not be forgotten that thirty-five years ago Haldane, Wigglesworth and Woodrow (1924), experimenting with the effects of ammonium chloride feeding in man, remarked on the acidosis with its concomitant loss of cell phosphate and potassium, and stated explicitly that in prolonged acidosis, phosphate and potassium depletion might have serious effects.

Many other conditions may be associated with hypokalaemia and all of them require that therapy be controlled with flame photometry. Among these are severe and prolonged vomiting or diarrhoea from any cause, and it is well to remember that continued gastric aspiration with or without lavage may lead not only to alkalosis and hypochloraemia, but also to hypokalaemia. So, too, in intestinal ileus, the coils of intestine may be greatly distended with fluid rich in salts, so that sodium and potassium deficiency may occur without any vomiting or diarrhoea at all. Hypokalaemia may also occur from the prolonged exudation of burns, in treatment with cortisone and ACTH, and it is sometimes induced along with acidosis when base-binding resins are given orally to restrict sodium intake in cases of oedema or high blood pressure. On the other hand, hyperkalaemia may be equally serious. It occurs in association with shock and secondary renal damage after crush injury, and it may be delayed by the application of a tourniquet (where this is feasible). Hyperkalaemia may also occur after copious blood transfusion with stored blood, and hence is not unknown in the replacement transfusion of erythroblastosis foetalis. It is frequent during the anuria of incompatible blood transfusion, but it may be succeeded by hypokalaemia if the patient recovers and copious diuresis supervenes. Among other causes of hyperkalaemia are deficiency of adrenal cortical hormones or of cortico-trophic hormone. There are other causes of potassium deficiency and excess besides those already mentioned, but enough has been said on this topic to show how flame-photometry has made us

blood urea is not high and plasma inorganic phosphate so far from being raised, is low.

Biochemical investigation has suggested that in this group, the several members of which show clinical differences, various types of defect exist for resorption by the renal tubules. Some of these are shown in Table 2 abridged from a table by Professor C. E. Dent; it embodies the work of Albright, Fanconi and of Dent himself (for references see Dent, 1952; 1956). In type 1, there is a failure in tubular resorption of the inorganic phosphate which was originally present in the glomerular filtrate, hence plasma inorganic phosphate falls. This in turn leads to liberation of phosphate from calcium phosphate in bone, and at the same time calcium is liberated leading to decalcification. It is possible that interpreting these findings on the basis of the solubility product of calcium and phosphate in plasma involves oversimplification, but if one assumes that plasma is saturated with regard to calcium phosphate, then one may write: $Ca \times phosphate = k$ (or more properly taking account of the valencies, $[Ca^{++}]^2 \times [PO_4^{\equiv}]^3 = k$). Clearly, if phosphate falls owing to failure of resorption, the plasma will tend to become unsaturated, and either phosphate must increase to normal at the expense of the calcium phosphate of bone or calcium must rise also at the expense of bone. Thus in either case calcium and phosphate must be liberated from bone. That in fact plasma calcium does not rise, must result from its excretion by the kidney. It will be noted that in some cases of this type there is also a failure of glucose resorption leading to glycosuria.

In the next group the defects are multiple. Failure of phosphate resorption again leads to decalcification, glycosuria is present, aminoacids are incompletely resorbed by the tubules and may be detected in the urine by chromatography. There is also a partial failure of water resorption with polyuria, and in some cases there is a failure in resorption of potassium, which may be associated with acidosis and potassium deficiency as in nephrocalcinosis. Cases of this type often with Milkman's fractures of bone have been called by the term 'Fanconi's syndrome'; the disease is familial. It may be noted that the cirrhosis which may occur in this condition and the cystine

TABLE 2. Findings in rickets (or osteomalacia) resistant to vitamin D

Type	Tubule resorption defect					Disease or syndrome
	Inorganic phosphate	Glucose	Amino acids	Water	Potassium	
1	Yes	No	No	No	No	Resistant rickets, idiopathic osteomalacia, Milkman's
2	Yes	Yes	No	No	No	
3	Yes	Yes	Yes	Yes	No	Fanconi, cystine rickets, cystinosis
4	Yes	Yes	Yes	Yes	Yes	
5	Yes	No	No	No	Yes	Hyperchoraemic nephrocalcinosis, Butler-Albright

In this group of diseases blood urea and non-protein nitrogen are normal, serum calcium normal or slightly low, plasma inorganic phosphate is low, plasma alkaline phosphatase usually high. Definite X-ray signs of decalcification are usually present. (Slightly modified from C. E. Dent, 1952.)

Tracer Techniques and the Synthesis of Haem

[illegible]

niques, is the synthesis of haem (for details and references, see Rimington, 1956). The tracer technique here involves the insertion of radioactive atoms into the haem molecule, and then by progressive degradation it is determined which fragment of the original haem molecule contains the tracer atom.

In 1946 Shemin and Rittenberg showed that feeding with glycine labelled with radioactive nitrogen quickly caused the haem of the haemoglobin molecule to become radioactive. No other aminoacid apart from serine had this effect. The globin moiety of the haemoglobin was not radioactive. This shows that haemoglobin lies outside the general metabolic pool of amino-acid interchange, for otherwise other radioactive amino acids would have quickly been incorporated into the globin moiety with a corresponding rise in radioactivity. Since, however, it is the haem which becomes radioactive after feeding with radioactive glycine, it follows that the glycine is used specifically in the synthesis of haem, and it may in fact be shown that the nitrogen is transferred to the nitrogen of the four pyrrole rings (Fig. 1). If carbon and not nitrogen in the glycine fed is radioactive, and the haem broken down by stepwise degradation, it is found that the carbon of the pyrrole rings and the carbon of the methene bridges becomes radioactive, but not the carbon of the methyl, vinyl and propionic side-chains. The latter, however, become radioactive on feeding with acetate containing ^{14}C. The acetate is probably introduced into the Krebs citric acid respiratory cycle and converted into a succinyl derivative. It seemed probable that a single compound combining the structures of glycine and acetate might be concerned in the synthesis of haem, and the demonstration of such a compound depended on researches which had been proceeding in other directions.

deposits which may be widely distributed in the cornea and elsewhere, have so far not been explained, and are subjects for further biochemical research.

Type 5 is associated with nephrocalcinosis, though calcification of the renal tubules is not an essential feature, but occurs at a late stage. Other features are the secretion of a neutral or slightly alkaline urine, with a *reduced* alkali reserve in the blood, a plasma inorganic phosphate which is low and a plasma chloride which is sometimes, but not always, raised. The chronic acidosis may be a factor in dehydration and potassium depletion, these in turn leading to lassitude and occasionally to hypokalaemic paralysis; decalcification of bones also occurs. The lesion is obviously complex. There is a failure in the tubular resorption of sodium and potassium, possibly due to lack of carbonic anhydrase in the tubule cells, capable of liberating hydrogen ions to exchange with sodium and potassium. Clearly if the glomerular filtrate is at pH 7 4 and the pH of the urine is 6, sodium and potassium ions must be resorbed by the tubules in exchange for hydrogen ions, and if hydrogen ions are not available, the exchange cannot occur and resorption of fixed base must fail. High urinary phosphate may be due in part to the acidosis, and also to defective absorption by the renal tubules, the latter explaining the low plasma phosphate. Low plasma phosphate in turn leads, as we have seen, to liberation of calcium from bone and a raised urinary calcium. The urine is thus rich in phosphate and calcium, and its neutral reaction is favourable to nephrocalcinosis and to the formation of phosphate calculi. The raised plasma chloride is presumably complementary to the decrease in the alkali reserve.

Definition of the type of case within this group of resistant rickets is of more than academic importance. Thus treatment of types 1 and 2 with massive doses of calciferol, controlled by estimations of serum calcium, results in relief of symptoms. Calciferol and possibly alkalis may also be useful in Fanconi's disease. In type 5 the hyperchloraemia and acidosis respond to treatment with alkali, though therapy must be continued indefinitely.

but it cannot be doubted that studies on these lines coupled with tracer studies of the various other constituents of the red cell will help to elucidate many of the factors concerned in the production of defective erythrocytes.

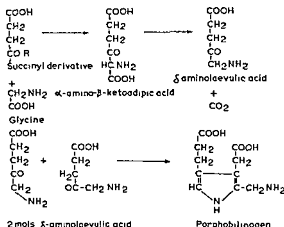


Fig 2. Scheme summarizing biosynthetic steps from glycine + succinate (from the tricarboxylic acid cycle) to porphobilinogen, the simplest pyrrolic substance known to be a precusor of haem and porphyrins

Congenital and Familial Metabolic Diseases

One of the most fruitful sources of investigation are the familial metabolic diseases. This was realized by Garrod over a century ago. There are many examples of such research. Thus Gibson (1948) has shown that congenital lack of a methaemoglobin reductase capable of reversing the conversion of oxyhaemoglobin to methaemoglobin, which occurs not only *in vitro* but also normally *in vivo*, is associated with methaemoglobinaemia. Again, the work of Holzel, Komrower and Schwarz (1957) has shown that in the galactosaemia of infants, an enzyme is lacking which converts galactose to glucose, or more properly galactose-l-phosphate to glucose-l-phosphate. The enzyme though mainly concentrated in the liver is presumably of widespread distribution, and is certainly present in red cells normally these glycolyse galactose far less efficiently than glucose, and in galactosaemia they are unable to ferment it at all In typical galactosaemia

In view of the fact that haem contains four pyrrole rings, it seemed likely that it might be synthesized from a compound containing a single pyrrole ring: no suitable compound, however, was known until Westall (1952) isolated a compound, porphobilinogen, from a patient with acute porphyria. Vast quantities of urine were required, and the substance was adsorbed on basic resin. Cookson and Rimington (1954) determined the constitution of this compound and showed that when incubated with a haemolysate of chicken erythrocytes, porphyrins were formed.

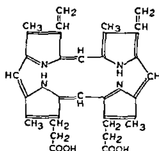


FIG. 1. Structure of protoporphyrin IX (The Fe of haem is attached to two of the nitrogen atoms).

In view of the fact that both glycine and acetate—the latter probably in the form of a succinyl derivative—were incorporated in the haem molecule, it was surmised by Shemin and Russell in the United States and by Neuberger and Scott in this country that delta aminolaevulic acid might replace both acetate and glycine in haem synthesis (Fig. 2). This was shown to be true for duck blood *in vitro* by Shemin and Russell (1953) and by Neuberger and Scott (1953). Then Dresel and Falk (1953) showed that when δ-aminolaevulic acid was incubated with a haemolysate of chicken red cells, porphobilinogen and porphyrins were formed in the absence of both glycine and acetate. It thus seems probable that δ-aminolaevulic acid is a precursor of haem, and Gibson, Neuberger and Scott (1955) have in fact demonstrated the existence of an enzyme which will cause the condensation of two molecules of δ-aminolaevulic acid to form one molecule of porphobilinogen. The story is not yet complete,

though they may seem quite healthy, sometimes manifest the characteristic metabolic defect in a minor symptomless form. Thus one or other parent of the galactosaemic infant may show an excessive rise in blood galactose during a galactose tolerance test; so, too, if phenylalanine is fed to the normal-seeming parent of a patient with phenyl-ketonuria, phenylpyruvate and other intermediates may appear in the urine, whereas in a truly normal individual all the phenylalanine would be converted to tyrosine. So, too, the parents and relatives of patients with hypophosphatasia, while free of all clinical abnormalities, may nevertheless excrete phosphoethanolamine in the urine. All these examples show how by chemical means genetic defects may be made manifest in heterozygous individuals, who to ordinary examination seem quite normal, and the possibility exists that by more refined and subtle methods, gene defects might become demonstrable in the commoner diseases where the familial influence is less clear: possible examples are obesity and arterial degeneration.

It was stated earlier that this little collection of examples should be regarded somewhat as an anthology, which could indeed be extended indefinitely. Thus investigation of the trace metals is likely to be very fruitful. The association of copper with hepatolenticular degeneration is a case in point (for references, see Bearn, 1957). Here the high copper content of the tissues has suggested treatment with such chelating agents as BAL (Cumings, 1948; Mandelbrote *et al.*, 1948) and penicillamine (Walshe, 1956): the latter, which may be given orally, has proved of especial value in promoting the excretion of copper in the urine and in effecting clinical improvement. Moreover, apart from these immediate gains, these various investigations have given a valuable 'pointer' to the importance of copper in cell metabolism.

The Future

Finally, what of further developments in the application of biochemistry to medicine? To some extent this will proceed on existing lines, examples of such advances are the demonstration that noradrenalin is raised in the blood in phaeochromocytoma

there are present debility, gastrointestinal upset, cirrhosis, cataract and mental deficiency, with early death. If, however, the disease is diagnosed soon enough, all these complications may be avoided by feeding the patient a diet in which galactose and lactose are absent.

In phenyl-ketonuria (for references see Knox and Hsia, 1957) there are severe mental deficiency, epilepsy, eczema and some lack of pigmentation. Phenylpyruvic acid and a number of allied compounds are found in the urine. A gene concerned with the development of an enzyme capable of oxidizing phenylalanine to tyrosine seems to be lacking. If very early in life such patients are fed on a diet deficient in phenylalanine, the chemical and physical abnormalities are much improved, and severe mental defect may be prevented. Yet another familial defect is hypophosphatasia (for references see Frazer, 1957). In this rare disease there is failure of proper calcification of the metaphyses, with deformity of a rickety type and a tendency to fractures. This is associated with a general lack of body alkaline phosphatase. *It is interesting to note that the urine of these patients* contains phosphoethanolamine (McCance, Morrison and Dent, 1955); possibly this is the natural phosphoric ester which when exposed to the phosphatase of bone liberates inorganic phosphate for calcification.

These familial metabolic disorders are clinically important from several points of view. First, the abnormal end products appearing in these patients may well be the metabolic intermediates of normal processes, which in disease fail to reach completion: phosphoethanolamine is perhaps a case in point. Second, these diseases pose the questions of why and how the accumulation of such intermediates leads to serious organic disease, and why failure to metabolize galactose or phenylalanine properly leads to physical deterioration and death. When the answer to these questions is known, we shall be better informed about the metabolism of the cell and the interaction of its enzyme systems. Thirdly, they throw light on genetics. In these familial disorders the homozygous patient has a genetic defect producing deficiency or absence of an enzyme system. The defect is inherited from the heterozygous parents, who

GIBSON, K. D., Neuberger, A. and SCOTT, J. J. (1955). *Biochem. J.* **61**, 618.
GIBSON, P. H. (1948). *Biochem. J.* **42**, 13.
HALDANE, J. B. S., WIGGLESWORTH, V. B. and WOODROW, C. E (1924). *Proc roy. Soc.* B, **96**, 1.
HOLZEL, A , KOMROWER, G. M. and SCHWARZ, V. (1957). *Am. J. Med.* **22**, 703
KNOX, W. E. and HSIA, D. Y.-Y. (1957). *Am. J. Med.* **22**, 687.
MCCANCE, R A., MORRISON, A. B. and DENT, C. E (1955). *Lancet*, **1**, 131.
MANDELBROTE, B. M , STANIER, B. M., THOMPSON, R. H. S. and THURSTON, M. N. (1948) *Brain*, **71**, 212.
NEUBERGER, A. and SCOTT, J. J. (1953). *Nature*, **172**, 1093
[illegible]
[illegible] *hem* , **166**, 621
[illegible] *em. Soc.* **75**, 4873
[illegible]
[illegible] 07–8.

(though the test is at present biological); also that 5-hydroxytryptamine is present in the blood and 5-hydroxyindole acetic acid (and similar compounds) in the urine of patients with argentaffinoma. So, too, it has been found that in the necrosis of various cells, a variety of enzymes are liberated, one or more of which may predominate according to the tissue involved: thus in coronary thrombosis and in liver damage a raised content of glutamic oxaloacetic transaminase may be demonstrated in the serum; in liver disease, too, plasma glutamic pyruvic transaminase is raised, while it is said that in pancreatic disease, and especially in carcinoma of that organ, there is an increase in serum leucine amino peptidase.

But it is probable that the future of biochemistry in relation to medicine rests not only on the study of the chemistry and physics of the body as a whole, or of tissues, or even of the individual cell as an integrated structure, but also, and perhaps more especially, on the study of the cell organelles: surface membrane, mitochondria, microsomes, nucleus and nuclear structures: this aspect of the medicine and pathology of the future has been stressed by Cameron (1954). In such studies radioactive tracers, chromatography, the electron microscope, genetic studies and also new techniques will all play a part. It is not too much to hope that ultimately the causes of the degenerative diseases and of cancer will be understood, and that effective treatment will become available.

REFERENCES

BEARN, A. G (1957) *Am J Med* **22**, 747.

BOYLE, R. (1725) *Philosophical Works*, ed. Peter Shaw, III, 616 Innys, Osborn & Longmans, London

[illegible] C. P. (1954) *Brit. med. J.* ii, 1061

57, 476

intriguing facts had emerged relating the adrenal cortex to a surprising variety of stresses (Sayers and Sayers, 1949). Though work was proceeding in animals, little was being done in man. Marthe Vogt (1943) had estimated the output from the adrenal vein in dogs by an exacting process of bioassay, and, much less well known, Eleanor Venning in Montreal (1945) had done a great deal of work on man by assaying the hormone content of urine by a complex process of bioassay involving glycogen deposition in the livers of adrenalectomized mice. She was able to show clearly two things. First, that the adrenal cortical hormone is normally present in human urine; secondly, that its content is increased by a variety of stresses such as surgical operations or by pregnancy. We were able to confirm that such an increase occurs in active disease states (Cope, Boysen and McCrae, 1951).

Chemical Assay

At this time the chemical nature of the hormone being estimated was not known with certainty, but it was recognized that the general group had a side-chain with reducing properties. The bioassay methods were so time-consuming that strenuous efforts were made to develop a chemical assay. At that time the special reaction for 17-ketosteroid estimation held almost undisputed sway as a measure of adrenal activity.

Two new general approaches to the chemical assay problem were soon made. The first was to estimate the so-called reducing steroids (Talbot, Saltzman, Wixom and Wolfe, 1945), and the second was to make use of the ability of the side-chain of these steroids to split off formaldehyde—permitting the estimation of the so-called formaldehydogenic steroids (Lowenstein, Corcoran and Page, 1946). (See Fig. 1.)

The difficulty with both these techniques lay essentially in lack of specificity and the lack of precise knowledge of what was being estimated. We now know that the actual proportion of real adrenal hormone in those extracts was often less than 10 per cent of the total reducing steroid. Metabolites and degradation products accounted for some more, but the total adrenal contribution was probably often not more than 50 per cent in ordinary urines. The rest was reducing substances largely unrelated

XVIII

The Measurement of Adrenal Activity in Man

C. L. COPE

THE scientific basis in medicine involves the approach by precise measurement, and in the realm of adrenal cortical physiology the advent of isotope-labelled steroids has made almost all kinds of measurement possible. They have in a sense called the bluff of all who seek precise methodology in this field. For now that so many measurements are available two new dilemmas are put before us. Firstly, the clinical or scientific value of the results obtained may be insufficiently great to justify the skill and labour used in gaining them. Secondly, now that we have the power to estimate or measure so many things, we need to be sure that we really know what we want to measure. We have indeed an *embarras de richesses*. That this stage has been reached in a mere ten years from a starting-point where only the most crude estimates of adrenal activity could be made in man is a good indication of the extremely rapid progress that can be made, when scientific interest is deeply stirred, as it was indeed by the discovery of the remarkable therapeutic effects of cortisone.

It is my proposal to describe some of my own travels through these ten years and to indicate to you the progress which has resulted from these technological advances.

Bioassay

My own interest began when, after return from the disturbing life of the War to pick up medicine again, I found some

normal quantity of reducing steroid in the daily urine by one method was two to six mg, with a mean of four mg. This could be increased four times or more by ACTH stimulation, but reduced adrenal function could not be detected with any certainty because even in severe Addison's disease, where adrenal function was almost zero, the content of reducing steroids in the urine might still be found not greatly reduced below the normal range (Cope and Bain, 1951).

This basal figure probably represented the amount of reducing contaminants unrelated to adrenal function. Such a test was unsuited to the detection of any but the most gross changes in adrenal function, and indeed often failed to show any reaction to severe medical stresses. Now Venning had already shown, and we had confirmed, that such incidents caused a rise in the biologically active hormone in the urine, so it seemed that the reducing steroids lacked the sensitivity of the biological test in showing activity change. Similar disadvantages lay with the formaldehydogenic steroids.

Further important advances resulted from two things. First, the demonstration by Schneider (1950) and by Zaffaroni, Burton and Keutmann (1950) that the active substances in urine were cortisone and cortisol, and by Bush and Sandberg (1953) that the active ones in blood were cortisol and corticosterone.

Secondly, the introduction by Porter and Silber (1950) of their chemical reaction for the particular side-chain which characterizes the adrenal type steroids (Fig. 1). The reaction gives a strictly linear relation between the colour developed and the amount of steroid when in pure solution and it is very sensitive. It was natural therefore to apply this to urine. Unfortunately complications immediately arose and it was soon apparent that there was no correlation between the colour obtained by this reagent from crude steroid extracts and the amount of hormone present as judged by bioassay (Cope and Hurlock, 1952). Indeed such was the inaccuracy that negative results were frequently obtained by the chemical method even in normal subjects.

to the adrenal. The same was true for the formaldehydogenic steroids. The result was that any change which may have occurred in adrenal steroid in the urine in response to change in adrenal activity was very heavily damped down, and largely

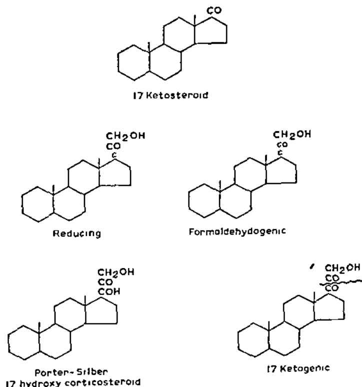


FIG 1. Types of steroid metabolite.

obscured, by the inert mass of contaminating substances. An additional problem with the formaldehydogenic steroids was the presence in the urine of some inhibiting factors to the liberation of the formaldehyde (Paterson and Marrian, 1953).

For such reasons, neither of these techniques ever really proved satisfactory. Since there was no certainty what was in fact being measured, the value of these reducing or formaldehydogenic steroid assays could only be assessed by observing the manner in which they coincided with clinical conclusions. The

glycol system is so much easier to use than most of the others that it is almost foolproof. When a simple chloroform extract of an acid urine is washed and separated on this system a wide variety of separate compounds is revealed but only 4 or 5 of these have reducing properties. If then the chromatogram is stained with blue tetrazolium, a reagent which forms a deep blue, water-insoluble, formazan when reduced, the position of these reducing steroids is clearly revealed. A rough estimate of the amount present can be obtained by direct visual comparison with standards treated in the same way, but a better method is to dissolve out the blue dye from the paper with acetone or preferably acid pyridine, and to estimate it in a colorimeter. Known standards are treated in the same manner and appropriate paper blanks are taken. The accuracy is not high, about ± 20 per cent in our hands, but this is more than sufficient for most clinical purposes (Cope and Hurlock, 1958). Two of these reducing spots are opaque to ultra-violet light showing they have Δ^{4-5} unsaturation. The upper of these is cortisol, the lower cortisone. In many hundred estimations of the upper or cortisol spot under a wide variety of clinical conditions we have never detected any reducing contaminant in this zone, although this is not true for the other reducing zones. The cortisone spot, for instance, is usually contaminated with other reducing steroids.

The ability to estimate the actual adrenal cortical hormone in urine relatively simply in this way is clearly of theoretical value but only empirical experience can decide whether it has clinical value. We will return to this point later.

Other steroid compounds which can be estimated in the same relatively simple manner are the first degradation products of cortisol metabolism, tetrahydrocortisone and tetrahydrocortisol, for both these have similar reducing properties. But 95 per cent of the total of these metabolites is present in the urine in the form of glucuronides so that a preliminary hydrolysis with a suitable β-glucuronidase enzyme is necessary before they can be extracted with chloroform. Except under very exceptional circumstances the presence of these substances in urine is proof that cortisol has been metabolized. If none has been given by mouth then the cortisol must have been formed in the adrenal gland.

CHROMATOGRAPHY

About this time the introduction of paper chromatography provided an extremely powerful tool with which to attack the problem better.

In our own laboratory we were able to show that even a simple paper chromatographic purification of the urine steroids permitted a subsequent chemical analysis to give a good correlation with the results of bioassay (Cope and Hurlock, 1952). A chemical method for the estimation of the actual hormone appeared then to be feasible, though the details would need further development. An alternative suggestion arose at about this time from the laboratories of Thorn. This was based on the fact that a very large proportion of the metabolites of adrenal steroids are eliminated in the urine in the form of glucuronides, and that these can be very effectively extracted from the urine by normal butanol. Reddy, Jenkins and Thorn (1952) developed a method based on the extraction of the urine with butanol, and the application to this extract of the Porter Silber reaction. Though this method was simple in theory, it proved, in our hands at least, highly unsatisfactory in practice. It was able to detect a gross increase in adrenal activity, but was quite unreliable for the detection of smaller changes, and Reddy himself (1954) when improving the method admitted that negative results were often obtained, because non-specific impurities interfered greatly with the reaction.

It is doubtful now whether this chemical assay problem can be overcome in urine without the use of some more complete purification procedure such as can readily be achieved by paper chromatography. There is a popular belief among many clinical chemists that paper chromatography of steroids is difficult, time-consuming, and unreliable in the separation of compounds. A very great deal depends on the solvent systems used. Generally speaking the less volatile the solvents the easier to manage are the chromatograms. The Bush type systems (Bush, 1951) rely essentially on volatile solvents and are somewhat fickle in use, requiring careful temperature control. The so-called Zaffaroni systems (Burton, Zaffaroni and Keutmann, 1951) are much less volatile and of these the toluene/propylene

significance of these different measures is seen in pregnancy, where in spite of a vast number of measurements made by one or other method (Table 1) doubt still persists as to whether or not adrenal function is increased (Martin and Mills, 1958; Migeon, Bertrand and Wall, 1957). We shall return to this point also.

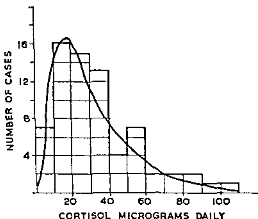


FIG 2. Urinary cortisol excretion Distribution in normal hospital subjects.

NORMAL VALUES OF INDIVIDUAL STEROIDS

Before we compare their relative virtues we must consider the normal range and values for some of the individual steroids. By our method the normal urine cortisol varies from a trace, i.e. less than 10 µg., to about 100 µg., with a mean of 43 µg. (Fig. 2). Tait and colleagues, using a highly accurate isotope method, have quite recently got similar figures (Jones, Lloyd-Jones *et al.*, 1959). The range for the tetrahydrocortisone and tetrahydrocortisol is wider, that for tetrahydrocortisone extending from a minimum of about 1,000 µg. to 3,500 µg. or more with a mean of 1,750 µg. daily (Cope and Hurlock, 1954).

It is instructive to compare the extent to which these measures change in the adrenal overactivity of Cushing's Syndrome. In a series of 12 cases of this disorder the mean rise in urine cortisol output was to 8·5 times the normal, whereas mean rise in

Before comparing the clinical value of estimations of these various urinary steroids there is one more group which must be mentioned. This is the so-called ketogenic steroids (Norymberski, Stubbs and West, 1953) which have rightly achieved a great clinical popularity (Fig. 1). The possible estimation of plasma cortisol as an index of adrenal activity is always a difficult procedure, and will not be considered here because its clinical possibilities have been excellently surveyed by Bayliss in an earlier Federation lecture (Bayliss, 1955).

The relative advantages of plasma and urine analyses may be compared in general terms. Plasma reflects more closely the internal environment, but its hormone concentration changes more rapidly from hour to hour. These concentrations are low, samples are of necessity restricted in volume and analyses are difficult. When completed they reflect conditions at a moment in time. In contrast urine samples are available in much larger volume, they contain larger quantities of the relevant steroids, their analysis is less difficult, and the results of such analyses reflect average conditions over the period of urine collections which may well exceed 24 hours. The disadvantages are that interpretation of urinary steroid results involves the assessment of variables in renal function and in steroid metabolism.

TABLE 1. Increases in some adrenal metabolites in late pregnancy urine

	% rise over non-pregnant
Free hormone (bioassay)	350
Formaldehydogenic steroids	100
Reducing steroids	100
17-ketosteroids	20-100
17-hydroxycorticosteroids	40
21-deoxyketols	700
17-ketogenic steroids	40
Plasma 17-hydroxycorticosteroids	200

These then are some possible measures of adrenal activity. It is remarkable how few attempts have been made to compare the relative values of these or of other measures of adrenal activity under different clinical circumstances. Yet an excellent indication of the urgent need for such a clarification in the relative

adrenal function will not be clearly revealed by this measure in many cases. But in such conditions of reduced adrenal activity, the tetrahydrocortisone is a far more satisfactory measure. For this the normal range is wide but the lower level of normal is rarely below 1,000 μg daily. Because the method for estimation of this metabolite is specific it is possible to estimate with reasonable clinical accuracy amounts of less than one-tenth of the lower normal limit, without any interference by the non-specific impurities which, in so many methods, prevent the measure of low degrees of adrenal function.

To illustrate the point by an analogy—one may judge the activity of a fire by its output of heat, but to determine whether or not the fire has gone completely out, it is often preferable to note whether smoke is still being produced.

The lowest tetrahydrocortisone outputs are seen in hypopituitarism or in total bilateral adrenalectomy. Less extremely low are the results in Addison's disease, where all degrees of depression of activity may be met according to the severity of the condition. In so far as adrenal activity cannot be maintained unless there is pituitary activity the presence of more than a trace of tetrahydrocortisone in the urine is incompatible with a marked degree of hypopituitarism. Thus we have in tetrahydrocortisone a sharply defined and specific test for hypopituitarism which has proved of value in confirming for instance the clinical results of pituitary implants in the treatment of inoperable carcinoma. Severe liver cirrhosis, though reducing tetrahydrocortisone output, does not interfere seriously with interpretation of the test (Fig. 4).

These tests can frequently be of great clinical value when other clinical data lead to ambiguous results. In such circumstances the demonstration of cortisol in the urine, or of tetrahydrocortisone, can provide effective proof of cortisol production. This has proved of great value, for instance, in advanced cases of pulmonary tubercle when weakness, hypochloremia, hyponatraemia and pigmentation may well simulate Addison's disease as a complication. They can be of value also in rare patients whose adrenals, though active, fail to respond to ACTH stimulation.

The relative behaviour and sensitivity of some of these

ketogenic steroids was only 3·8 times normal, that of tetrahydrocortisone only 1·9 times and of 17-ketosteroids only 1·4 times normal. The same essential changes may be seen in the response to ACTH stimulation. Here again the rise in urine cortisol exceeds

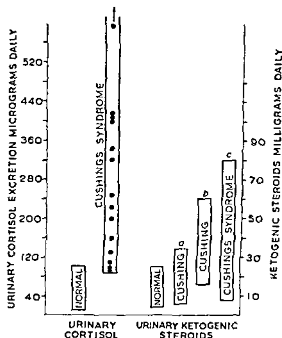


FIG. 3. Comparison of urinary cortisol and of urinary ketogenic steroids in diagnosis of Cushing's Syndrome

that of any other measure by a similar amount. That this sensitivity may be of clinical diagnostic value in the diagnosis of adrenal overaction of Cushing's type is shown in Figure 3 which compares the normal and Cushing's ranges for urinary cortisol and for its next competitor, the ketogenic steroids. The overlap between normal and abnormal in this series at least is almost zero. Urinary cortisol is thus revealed as the most sensitive indicator of increased adrenal activity and has considerable diagnostic value.

But we have seen that the normal urinary cortisol output may be as low as a trace only. It follows therefore that reduced

measures of adrenal cortical activity can be very well seen when attempts are made to inhibit the enhanced adrenal function of Cushing's Syndrome, a test which we believe to have some clinical value.

Figure 5 shows that urinary cortisol falls quickest and most completely. The falls in the metabolites are also large but more sluggish, no doubt because their turnover rates are slower than that of cortisol. Ketogenic steroids and ketosteroids, however, reveal themselves as very much less sensitive and more sluggish than any of the other tests for showing these changes.

Purely clinical comparison thus reveals the relative value of these reasonably simple tests, but it does not tell us how accurately they relate with the actual changes in cortisol production. Nor indeed was it possible, except in rare cases of cannulation of the adrenal vein at operation, to determine cortisol production rate until the advent of isotopic tracers.

Use of Radioactive Cortisol

But it is the availability of these isotope-labelled steroids which now enables us to measure more accurately so many things.

The estimation of plasma cortisol, for instance, present in only 1/5,000 the concentration of resting glucose, has always been difficult but it is greatly simplified by the addition of a minute trace of the isotopic steroid (Bondy and Upton, 1957).

The first useful direct estimations of actual cortisol production rate in man were made by Peterson and Wyngaarden (1956) using classical principles. In their method a small dose of ^{14}C-labelled cortisol is injected intravenously into the blood stream where it gets progressively diluted with the non-radioactive cortisol produced by the adrenal cortex. By taking serial blood samples and determining the ^{14}C content of the cortisol in each sample, the rate of this dilution is determined and the production of endogenous cortisol which is needed to effect this change can readily be calculated.

While the method produces valid results it suffers from the disadvantage that several skilled blood analyses are required at intervals and that the period of observation is of necessity short.

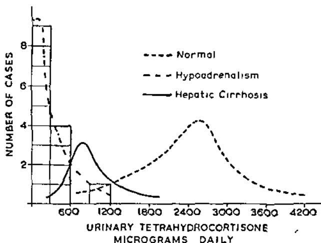


FIG. 4. Tetrahydrocortisone excretion in normal subjects, in hepatic cirrhosis and in hypopituitarism.

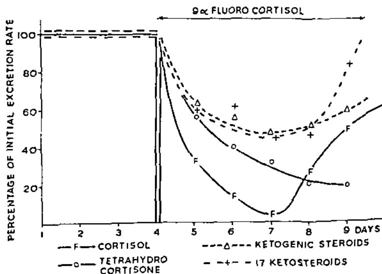


FIG 5. Urinary steroid excretion as indices of adrenal inhibition.

Tetrahydrocortisone, however, appears to fulfil these needs admirably. It is present in relatively large amount in urine—usually from 1 to 4 mg. daily—it can be separated with sufficient purity for these purposes in a single chromatographic step, and, except under rare circumstances, it is not derived in appreciable quantity from steroids other than cortisol.

If a chloroform extract of a hydrolysed urine is obtained from a patient who has been given a trace dose of ^{14}C cortisol, and this extract is submitted to a single-stage chromatography, the tetrahydrocortisone is largely separated from other steroids and contaminants, but because of the immense variety of steroids and impurities likely to be found in any urine complete purity would be most unlikely. Because of the dose of radioactive cortisol given, all adrenal metabolites of cortisol known and unknown will be labelled with this ^{14}C, although of course (owing to dilution with the endogenous supply) at a much lower specific activity than the original dose. The metabolite tetrahydrocortisone is a reducing steroid and reduces the tetrazolium reagent. If then the reducing steroid content of the chromatogram is compared cm. by cm. with the ^{14}C content, a good coincidence between the two will show that all the reducing steroid at this point on the chromatogram is derived from cortisol metabolism. Such a coincidence is found to occur. It does not strictly prove that the steroid in this zone is tetrahydrocortisone, although this is highly probable since similar coincidence also occurs on other chromatographic systems. But knowledge of the nature of the metabolite is not really necessary for our purpose. We are concerned only to find a suitable reducing steroid which is derived from cortisol metabolism and which can readily be freed from disturbing contamination. This the tetrahydrocortisone, or tetrahydrocortisol spot, provides adequately after a single paper chromatography on the easiest of systems, toluene/propyleneglycol. If then one elutes this spot with ethanol it is a quite simple procedure to divide it into two equal parts and determine the reducing steroid in one half and the ^{14}C in the other. This is the only stage in the whole analysis of the urine sample which needs to be done with quantitative accuracy. The two measurements give the specific activity of

The result is a measure of the mean cortisol production rate over a period of 3 to 4 hours. Using this method Peterson and Wyngaarden obtained figures of 17 mg. to 29 mg. daily with a mean of 21 mg.

Whilst such a measurement can be of immense clinical value for some purposes, a simpler procedure which gives a mean of the daily output would also have very great clinical advantages. We have devised such a method which makes relatively small demands on technical skill and should be within the powers of any competent clinical laboratory with isotope facilities, though it is scarcely suitable as yet for routine clinical use (Cope and Black, 1958a).

In this method a known dose of ^{14}C-cortisol in water without added carrier is given by mouth. Most of it is very rapidly absorbed into the blood stream where it mixes with the normal cortisol pool. There it is metabolized in the same way as the endogenous cortisol, and the products of metabolism are excreted in the urine, all labelled with ^{14}C in the same proportion. The recovery of most of the ^{14}C from the urine is proof that the dose was absorbed in that patient, and also proof that metabolite excretion is prompt. If now a suitable metabolite of cortisol metabolism is isolated from the urine and its specific activity or ^{14}C content per μg. steroid is determined, then by a simple arithmetical calculation the daily cortisol production needed to effect this degree of dilution is easily obtained. The principle is the well-known one of isotope dilution. It would be more reliable to inject the ^{14}C cortisol intravenously but that introduces undesirable complications and is not usually needed except for special purposes.

Practice is of course often less simple than theory, and such a method necessitates the selection of a metabolite derived only from cortisol and the ability to extract it and purify it sufficiently to determine the specific activity or ^{14}C content. It is a peculiarity of this type of isotope procedure that no analytical accuracy is required until the final stages, for a quantitative recovery is not needed at any stage. Urine cortisol itself should be suitable for determining the degree of isotope dilution but the quantity is often small and its analysis is technically difficult.

Tetrahydrocortisone, however, appears to fulfil these needs admirably. It is present in relatively large amount in urine—usually from 1 to 4 mg. daily—it can be separated with sufficient purity for these purposes in a single chromatographic step, and, except under rare circumstances, it is not derived in appreciable quantity from steroids other than cortisol.

If a chloroform extract of a hydrolysed urine is obtained from a patient who has been given a trace dose of ^{14}C cortisol, and this extract is submitted to a single-stage chromatography, the tetrahydrocortisone is largely separated from other steroids and contaminants, but because of the immense variety of steroids and impurities likely to be found in any urine complete purity would be most unlikely. Because of the dose of radioactive cortisol given, all adrenal metabolites of cortisol known and unknown will be labelled with this ^{14}C, although of course (owing to dilution with the endogenous supply) at a much lower specific activity than the original dose. The metabolite tetrahydrocortisone is a reducing steroid and reduces the tetrazolium reagent. If then the reducing steroid content of the chromatogram is compared cm. by cm. with the ^{14}C content, a good coincidence between the two will show that all the reducing steroid at this point on the chromatogram is derived from cortisol metabolism. Such a coincidence is found to occur. It does not strictly prove that the steroid in this zone is tetrahydrocortisone, although this is highly probable since similar coincidence also occurs on other chromatographic systems. But knowledge of the nature of the metabolite is not really necessary for our purpose. We are concerned only to find a suitable reducing steroid which is derived from cortisol metabolism and which can readily be freed from disturbing contamination. This the tetrahydrocortisone, or tetrahydrocortisol spot, provides adequately after a single paper chromatography on the easiest of systems, toluene/propyleneglycol. If then one elutes this spot with ethanol it is a quite simple procedure to divide it into two equal parts and determine the reducing steroid in one half and the ^{14}C in the other. This is the only stage in the whole analysis of the urine sample which needs to be done with quantitative accuracy. The two measurements give the specific activity of

the metabolite. From this the cortisol production rate is very readily derived in the following manner:

Let us suppose that D is the dose of ^{14}C given, and that C_u is the total ^{14}C recovered in the urine in the subsequent 24 hours. If C_u equals D then one can be quite certain that the whole dose has been given, that it was completely absorbed and that it has been fully excreted. But, in practice, complete recovery of the ^{14}C is not obtained. An ambiguity in interpretation at once arises, for it may be on the one hand that the whole dose was absorbed but that only 80 per cent is excreted in the urine, or alternatively that 20 per cent of the urine may have been lost. The other possibility is that only 80 per cent of the dose was actually absorbed but that this was completely excreted. In the first case an 80 per cent excretion of adrenal metabolites is indicated, in the latter a 100 per cent efficiency. The truth will lie between these two extremes and two estimates of the excretion efficiency may be made.

Now, there is no reason to expect that the radioactive cortisol will behave in a manner different from natural or endogenous cortisol. The ^{14}C will therefore be distributed evenly among all the metabolites of cortisol. The specific activity of all these should be similar so that, once the ^{14}C content per microgram of a representative metabolite has been found, the total mass of metabolites, known and unknown, can be calculated, since it is the total ^{14}C content of the urine sample divided by the specific activity (C_u/SA_u).

Since we now know the excretion efficiency for the metabolites of cortisol and the total mass of metabolites, the latter divided by the former represents the actual daily production of cortisol This is either

$$C_u/SA_u \text{ or else } D/C_u \times C_u/SA_u = D/SA_u.$$

The difference between these two estimates indicates the degree of reliability in that particular test. This method has the immense advantage of checking automatically all ward errors such as loss of part of the dose or part of the urine collection.

It may be objected that the oral route is not a representative one because the steroid dose must pass through the liver and so

risk destruction before it enters the general metabolism. Whilst this is a sound theoretical objection, it does not seem to be valid in practice for approximately the same result is obtained in a given individual, if oral and intravenous routes are compared on different days.

The mean normal resting hydrocortisone production rate determined in this way is in man 13·5 mg. with a range of 5 to 28 mg. daily, and in young women 10·5 mg. with a range of 5 to 24 mg. (Table 2).

TABLE 2. Cortisol production rates
mg. daily

Normal—male (12) 5 to 28	mean 13 5
female (8) 5 7 to 24	mean 10 5
[illegible]	
[illegible]	mean 0 9
[illegible]	mean 8 0
[illegible]	mean 25
Late pregnancy (8) 15 to 40	mean 25
Cushing's Syndrome (3) 20 to 60	mean 45

(Figures in brackets are number of cases in each group)

These figures are considerably lower than some estimates that have been made, and are, in fact, lower than the results of Peterson and Wyngaarden (1956) obtained by their different method of isotope dilution. The fact that our results are lower is of some importance. For one possible objection to the method is that some of the tetrahydrocortisone analysed may in fact be derived from sources other than cortisol metabolism. Any contamination in this manner would artificially *raise* the apparent production rate, so that our very low values help to give us confidence that this is not an appreciable complicating factor.

A probable reason for the [illegible] Peterson an[illegible] morning ho[illegible] mal. Our own results represent a mean for the 24 hours and will include the period of sleep when output is likely to be minimal. It is of considerable interest that a single measurement made by Peterson and Wyngaarden at night gave a figure of 12 mg., a result essentially the same as our own mean for males.

the metabolite. From this the cortisol production rate is very readily derived in the following manner:

Let us suppose that D is the dose of ^{14}C given, and that C_u is the total ^{14}C recovered in the urine in the subsequent 24 hours. If C_u equals D then one can be quite certain that the whole dose has been given, that it was completely absorbed and that it has been fully excreted. But, in practice, complete recovery of the ^{14}C is not obtained. An ambiguity in interpretation at once arises, for it may be on the one hand that the whole dose was absorbed but that only 80 per cent is excreted in the urine, or alternatively that 20 per cent of the urine may have been lost. The other possibility is that only 80 per cent of the dose was actually absorbed but that this was completely excreted. In the first case an 80 per cent excretion of adrenal metabolites is indicated, in the latter a 100 per cent efficiency. The truth will lie between these two extremes and two estimates of the excretion efficiency may be made.

Now, there is no reason to expect that the radioactive cortisol will behave in a manner different from natural or endogenous cortisol. The ^{14}C will therefore be distributed evenly among all the metabolites of cortisol. The specific activity of all these should be similar so that, once the ^{14}C content per microgram of a representative metabolite has been found, the total mass of metabolites, known and unknown, can be calculated, since it is the total ^{14}C content of the urine sample divided by the specific activity (C_u/SA_u).

Since we now know the excretion efficiency for the metabolites of cortisol and the total mass of metabolites, the latter divided by the former represents the actual daily production of cortisol. This is either

$$C_u/SA_u \text{ or else } D/C_u \times C_u/SA_u = D/SA_u.$$

The difference between these two estimates indicates the degree of reliability in that particular test. This method has the immense advantage of checking automatically all ward errors such as loss of part of the dose or part of the urine collection.

It may be objected that the oral route is not a representative one because the steroid dose must pass through the liver and so

and patient collaboration in each individual case. There is a good prospect of radioactive cortisol becoming more widely available and radioactive cortisone has just been put on the market by the Radiochemical Centre, Amersham.

Some Clinical Applications

Clearly such a method has a wide clinical application. It enables us, for instance, to help solve the vexing problem of whether or not cortisol production is raised in pregnancy. This can be done without risk to the mother or to the foetus from radiation because elimination of the steroid is rapid and the dose is small. We have investigated this problem in a small series of patients with the following results. In a control group of 8 non-pregnant women of between 20 and 35 years the mean production rate was 10·5 mg. daily. In a comparable group of 8 women in the last month of pregnancy, the mean was 25 mg. daily. Thus the production is on average increased to 2½ times the non-pregnant level by the last month of pregnancy (Cope and Black, 1959). It is of interest that Tait and colleagues have recently measured an approximately threefold increase in aldosterone secretion also in late pregnancy (Jones, Lloyd-Jones *et al.*, 1959).

How does such a degree of activity compare with that encountered under other conditions?

We have seen that a 10 or 20 times rise may occur after ACTH stimulation and that in Addison's disease a fall of below one-tenth (1 mg.) may result.

In general medical disease, the rise is relatively small, not as great as in pregnancy, but in hyperthyroidism a greater stimulation seems to occur and outputs are comparable to those seen in late pregnancy.

It is relatively easy with this method to measure the inhibition of cortisol production which results from administration of similar steroids. For the steroid metabolite is a specific one and if care is taken its analysis is not interfered with by metabolites of the alternative synthetic steroid. With prednisone, for instance, the cortisol output is inhibited to about 2 mg. daily. We have failed to get a lower inhibition than that and the

It is interesting that cortisone acetate can be used in the same manner and for the same purpose to measure cortisol production. This is possible because it is largely turned into cortisol in the body. The normal value for cortisol production obtained with this steroid is precisely the same as when cortisol is used. But absorption from the stomach is slightly less efficient resulting in a rather bigger range between the two alternatives.

It is of interest to see what happens when normal adrenal cortical function is disturbed (Table 2). Under the stimulus of ACTH therapy rise in output is very variable but it may be as much as 250/mg. daily. At the other extreme, damage to the adrenal cortex as in Addison's disease, and its atrophy due to hypopituitarism, both result in a fall in output to below 1 mg. daily. An interesting point arises here which is unusual in endocrine assays. It is common experience in most assay methods that although one may be able to estimate hormones or metabolites when they are present in excess, it is usually impossible to detect them when they fall to the lower limits of the normal range, because they fall below the limit of sensitivity of the methods used and are lost in the inevitable impurities and interfering substances. This is not true, however, with the isotope method for cortisol production, and its percentage reliability is just as high at very low outputs as at high ones because the amounts of ^{14}C measured are not affected by the adrenal activity and the reduced content of tetrahydrocortisone in the urine in hypoadrenalism is easily compensated for by taking a larger sample of urine for analysis.

The validity of such a method is dependent on a lack of any significant recirculation of the ^{14}C within the body. The excretion curve for this ^{14}C and the behaviour of injected metabolites both support the view that no such recirculation occurs and that once cortisol has been broken down the degradation products are treated as foreign bodies.

We thus have a relatively simple method of measuring the daily cortisol production which is readily applicable to nearly all hospital patients and which involves an almost negligible exposure to radiation. It has the enormous advantage also of providing an automatic internal check on the efficiency of ward

the latter from the former. An excretion of 27 mg. 17 ketosteroids daily for instance may be associated with cortisol outputs of from 10 to 200 mg. daily.

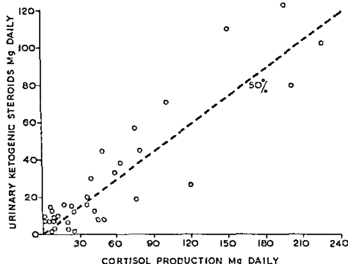


FIG. 7 Relation of urinary ketogenic steroid excretion to actual cortisol production

The urinary cortisol level, which is so sensitive for some purposes, also shows a rather poor correlation with the actual production rate. The cortisol in the urine represents from 0·1 to 1 per cent of the amount produced. This variability may well be due to the fact that a large part of the plasma cortisol is bound to protein and the amount so bound is likely to vary in different individuals.

The excretion of the first-stage metabolites, tetrahydrocortisol and tetrahydrocortisone, taken either separately or together, shows a much better relation in the middle and lower ranges of activity, and their excretion represents from 10 to 15 per cent of production. But in the higher ranges of activity the correlation seems to be less good and the proportion of total production excreted as these tetrahydro compounds is reduced.

The urinary 17-ketogenic steroids of Norymberski show the

reason may well be the existence of a slight artefact here, for Vermeulen (1958), using radioactive prednisolone, has shown that a small amount of tetrahydrocortisol is found among its metabolites. The creation of this additional fraction will raise

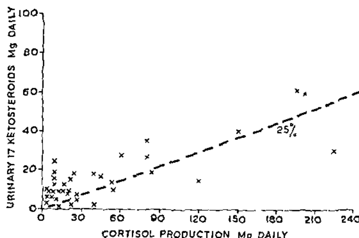


FIG. 6. Relation of urinary 17-ketosteroid excretion *to actual cortisol production*

the apparent production rate slightly above its true value. Under most circumstances this source of error seems to be a minor factor, but very occasionally it can be a factor causing erroneous conclusions to arise. This may occasionally happen in adrenal congenital hyperplasia in which condition hydrocortisone production is impaired, but in which nevertheless considerable amounts of tetrahydrocortisone may be formed, some very probably from other parent steroids not at present recognized (Cope, 1959).

With a method of this type it is for the first time feasible to compare the results of the more routine methods of assaying clinically adrenal function with the actual production rates.

We have already seen that 17-ketosteroids are a very poor and insensitive indicator of changing adrenal cortical function and this is well revealed by such a comparison (Fig. 6).

If urinary 17-ketosteroids are plotted against cortisol production the scatter is such that no conclusions can be drawn about

CONCLUSIONS

What general conclusions can we draw from all this? I suggest that the main one is that too little effort is yet made in clinical studies of adrenal activity in man to take the elementary precaution of adapting the tool to the particular job in hand. There are now widely varied methods of measurement of adrenal activity available. Some are accurate but exacting to perform, but some others are relatively simple though yielding precise information which surely is the essence of the scientific approach. Much greater efforts could and should be made to choose the type of adrenal function test best suited to the particular job in hand.

The introduction of isotope-labelled steroids makes possible much better reference standards by which to judge the reliability of more simple tests. No longer are we compelled to take the view that the criterion of a good test is that its results agree with the clinical impression. We can at last express the activity of the adrenal cortex in absolute terms and that is a step forward indeed in the scientific investigation of this gland and of the study of steroid metabolism in man.

REFERENCES

BAYLISS, R. I S (1955) *Brit med J* i, 495.
BONDY, P K and UPTON, G. V (1957) *Proc Soc. exp Biol (N Y)* 94, 585.
BURTON, R B, ZAFFARONI, A and KEUTMANN, E H (1951) *J biol. Chem* 188, 763.
BUSH, I E (1951) *Biochem J.* 50, 370
BUSH, I E and SANDBERG, A. A (1953) *J biol Chem.* 205, 783
COPE, C L (1959). *Brit med J* i, 815.
COPE, C L. and BAIN, M (1951) *Clin. Sci* 10, 161.
COPE, C L. and BLACK, E G (1958a) *Clin Sci* 17, 147
COPE, C L and BLACK, E G (1958b) *Brit. med J* i, 1020.
COPE, C L and BLACK, E G (1959). *J Obst Gynec Brit Empire* 76, 404.
COPE, C L, BOYSEN, X and MCCRAE, S. (1951). *Brit med J* ii, 762
COPE, C L and HURLOCK, B (1952) *Brit med J* ii, 1020
COPE, C L and HURLOCK, B (1953) *Med Soc. Endocrin* 2, 25
COPE, C L. and HURLOCK, B (1954). *Clin Sci* 13, 69

best general correlation with actual production rate. From Figure 7 it can be seen that they represent on average about 50 per cent of the total production. That so large a proportion of the metabolites is represented by this group of steroids is important, for it may be taken as a general rule that the bigger the percentage of the total metabolites which can be estimated by *any analytical procedure, the more closely will the result correlate with the actual secretion rate.*

That is why the ^{14}C-cortisol method is so reliable, because, in effect, it measures the total mass of metabolites.

There are other adrenal activities known and probably still more to be discovered. Cleverer groups than mine have applied essentially similar methods to the study of aldosterone secretion. The production rate for this steroid is only about one-hundredth of the amount of cortisol production so that its study demands much more refined methods.

Nevertheless, two groups have successfully tackled the problem. As with cortisol, three possible methods theoretically exist. The first, the measurement of the rate of dilution of a sample of radioactive aldosterone introduced into the blood stream, is not technically feasible because of the excessively small quantities involved. In the other two we have the choice, as with cortisol, of measuring the specific activity either of the urine aldosterone, or of a metabolite. Both methods are difficult, and use of the metabolite has not got the great advantage it has with cortisol, because the gain in quantity of steroid handled is much less.

Jones, Lloyd-Jones, Riondel, Tait and their colleagues (1959) in the United Kingdom have used the aldosterone and Lieberman and his team at Columbia, New York, the metabolite (Ulick, Laragh and Lieberman, 1958). The results of both groups agree. Production is from 50 to 300 μg. daily on ordinary diets, but may rise over 1,000 μg. in nephrosis, or on low salt intake. There is growing evidence, too, that, as with urine cortisol, the relation between urine aldosterone and aldosterone secretion is a relatively poor one. But aldosterone techniques are scarcely available yet except to a few enthusiasts.

best general correlation with actual production rate. From Figure 7 it can be seen that they represent on average about 50 per cent of the total production. That so large a proportion of the metabolites is represented by this group of steroids is important, for it may be taken as a general rule that the bigger the percentage of the total metabolites which can be estimated by any analytical procedure, the more closely will the result correlate with the actual secretion rate.

That is why the ^{14}C-cortisol method is so reliable, because, in effect, it measures the total mass of metabolites.

There are other adrenal activities known and probably still more to be discovered. Cleverer groups than mine have applied essentially similar methods to the study of aldosterone secretion. The production rate for this steroid is only about one-hundredth of the amount of cortisol production so that its study demands much more refined methods.

Nevertheless, two groups have successfully tackled the problem. As with cortisol, three possible methods theoretically exist. The first, the measurement of the rate of dilution of a sample of radioactive aldosterone introduced into the blood stream, is not technically feasible because of the excessively small quantities involved. In the other two we have the choice, as with cortisol, of measuring the specific activity either of the urine aldosterone, or of a metabolite. Both methods are difficult, and use of the metabolite has not got the great advantage it has with cortisol, because the gain in quantity of steroid handled is much less.

Jones, Lloyd-Jones, Riondel, Tait and their colleagues (1959) in the United Kingdom have used the aldosterone and Lieberman and his team at Columbia, New York, the metabolite (Ulick, Laragh and Lieberman, 1958). The results of both groups agree. Production is from 50 to 300 µg. daily on ordinary diets, but may rise over 1,000 µg. in nephrosis, or on low salt intake. There is growing evidence, too, that, as with urine cortisol, the relation between urine aldosterone and aldosterone secretion is a relatively poor one. But aldosterone techniques are scarcely available yet except to a few enthusiasts.

XIX

Some Aspects of Iron Metabolism

D. F. CAPPELL

IRON metabolism is so vast a subject that only certain aspects can be touched upon in this lecture, and I shall deal chiefly with matters that concern the pathologist and shall try to illumine them from my own experience in experimental and human pathology. There has been a certain continuity of interest in iron in the Pathology Department at Glasgow beginning with the work of Robert Muir on the marrow, his fundamental studies on haemolysis *in vitro* with Browning and others, and *in vivo*, along with Shaw Dunn and McNee. McNee's work on haemolytic icterus thus began in Glasgow and was continued in Freiburg where it contributed to Aschoff's concept of the reticulo-endothelial system. Soon thereafter I became involved in the problems of vital staining. I used saccharated oxide of iron, at first merely as an inert substance to demarcate the reticulo-endothelial system and later to produce so-called reticulo-endothelial blockade. Soon, however, it was observed that this substance was far from inert and experiments were then conducted to follow the fate of the iron and to assess whether it was harmful when administered in excess. Interest was, of course, stimulated by Muir and Dunn's observations on the iron content of the organs in haemochromatosis and attempts were made almost simultaneously by Polson in Manchester and by me in Glasgow to reproduce this disease by excessive parenteral administration of iron.

Iron is a remarkable substance. The human body contains on an average about 4–5 g. of iron, of which 2–3 g. is incorporated in haemoglobin, about 1–1·5 g. as storage iron, and most of the

HEARD, R. D. H., SOBEL, H. and VENNING, E. H. (1946). *J. biol. Chem.* **165**, 687 and 699.

JONES, K. M., LLOYD-JONES, R., RIONDEL, A., TAIT, J. F., TAIT, S. A. S., BULBROOK, R. D. and GREENWOOD, F. C. (1959). *Acta Endocrin.* **30**, 321.

LOWENSTEIN, B. E., CORCORAN, A. C. and PAGE, I. H. (1946). *Endocrinology*, **39**, 82.

MARTIN, J. D. and MILLS, I. H. (1958). *Clin. Sci.* **17**, 137.

MIGEON, C. J., BERTRAND, J. and WALL, P. E. (1957). *J. clin. Invest.* **36**, 1350

[illegible]

SAYERS, G. and SAYERS, M. A. (1949). 'The Adrenal Cortex.' *Ann. N.Y. Acad. Sci.* **50**, 522.

SCHNEIDER, J. J. (1950). *J. biol. Chem.* **183**, 365.

TALBOT, N. B., SALTZMAN, A. H., WIXOM, R. L. and WOLFE, J. K. (1945). *J. biol. Chem.* **160**, 535.

ULICK, S., LARAGH, J. H. and LIEBERMAN, S. (1958). *Trans. Ass. Amer. Phycns.* **71**, 225

[illegible]

facts and figures about these are given in Table 1. Since our paper on this subject two years ago we have had the opportunity of analysing some additional cases and I shall refer to these later.

I shall begin with the proposition that the long-continued presence of excessive deposits of iron in the tissues does not *necessarily* bring about such distinctly harmful effects on the parenchymal cells as to lead *inevitably* to destruction and fibrosis of parenchymatous organs such as liver and pancreas. I do not claim that excessive iron storage from transfusion is never associated with signs and even symptoms of organ damage, but I think it would be fair to insist that if the storage of iron from transfused blood is to be held responsible for hepatic and pancreatic fibrosis then the severity of the damage to these parenchymatous organs should bear a well-defined relationship to the quantity of iron stored in the organs and the length of time it has been there. In my view, analysis of cases reported in the literature as 'transfusional' or 'exogenous' haemochromatosis brings to light such wide discrepancies that there must be factors other than the quantity of iron to be taken into account.

My material is divided into two groups, who had been transfused in varying degree. I shall give a general account of our findings and illustrate my thesis especially by a more detailed consideration of Case No. 4 of Group I. The essential and relevant facts are given in Table 1 and the amount of blood transfused and the time of survival are adequate to give every opportunity for signs of damage to the organs to have occurred. The amount of iron stored in the organs also is very high, and is even greater than that found in haemochromatosis (Sheldon, 1935). Yet the microscopic examination of the tissues shows clearly that there is no real fibrosis or destruction despite the prolonged iron overload.

The general distribution of the iron pigment. The two Cases (4 and 5) which had received the largest amounts of blood showed brownish pigmentation of the skin, generally slight and most obvious in the legs. Pronounced rusty-brown pigmentation of the organs from haemosiderosis was present in all cases, the

remainder (about 0·3 g.) in the intracellular respiratory enzymes, the cytochromes, catalases etc., and a small amount is bound in the plasma (0·004 g.). In certain pathological conditions in man the iron content of the liver alone may be over 50 g. and the total body iron about 100 g., that is, twenty times the normal. There are few substances which can accumulate to this remarkable extent with so little obvious ill effect. It is important that apart from blood loss and loss by exfoliation of cells from skin and mucous surfaces, there is little excretion of iron once it has gained entrance to the tissues (McCance and Widdowson, 1937).

Conditions of Excessive Iron Storage

Gross excess of stored iron is found in the tissues in a variety of pathological states of diverse etiology. The chief of these are:

(*a*) idiopathic —haemochromatosis
(*b*) nutritional—Bantu siderosis
—Kaschin-Beck disease
(*c*) blood dyscrasia —haemolytic anaemia
—Addisonian anaemia
—thalassaemia
(*d*) iatrogenic —transfusional siderosis
—siderosis of prolonged iron administration

Of all the conditions of excess iron storage, that resulting from repeated blood transfusions is the best documented, since the total amount of blood administered is known, and this represents a minimum addition to the body's store at the rate of 0·5 g. iron for each litre of blood transfused. This is, of course, depleted by the usual sources of loss and by haemorrhage, which is often a feature of the condition for which transfusion has been required. Blood transfusion is today so easily available and is so much used to keep alive patients with aplastic and other anaemias unresponsive to other treatment that I make no apology for showing the later effects on the tissues of long-continued transfusion therapy. We have studied several cases in which a large number of transfusions had been given and some

liver, spleen, bone marrow, pancreas, stomach, choroid plexus, thyroid and adrenals being notably affected, but the kidneys were relatively free from iron. In Cases 4 and 5 chemical analyses revealed over 55 g. of iron in the livers and the sum of that in the other principal organs and tissues probably accounted for at least as much again in each case. The amount of iron was thus fully equal to that found in classical haemochromatosis.

Microscopic findings. The findings in the five cases are so similar that a general description will suffice; all the iron was in the ferric state.

Skin and subcutaneous tissues. In Cases 4 and 5, the skin of the abdomen and chest wall showed iron-storing macrophages most numerous around the sweat glands, the epithelial cells of which also show a distinct iron reaction of varying intensity in different parts of the skin.

The cells lining the ducts and acini of the *mammary glands* contain iron granules in the zone between nucleus and lumen (Plate XVIII, Fig. 1),[1] like that observed in the mammary epithelium of animals receiving intravenous saccharated oxide of iron (Cappell, 1929) and in the mammary epithelium in classical haemochromatosis (Cappell *et al*., 1957). Throughout the mammary stroma and adipose tissues generally, e.g. retroperitoneal etc., macrophages heavily laden with iron are numerous. In *gastric* mucosa (Plate XVIII, Fig. 2) pigmentation is maximal in the fundus and becomes less marked in the pyloric region. Iron is present in the deepest parts of the glands in both the chief and oxyntic cells in all except Case 2, but in Cases 4 and 5 the staining extends to involve the mucous neck cells but not the surface epithelium. Iron-positive phagocytes occur densely in the lamina propria, in small numbers between the muscle bundles but numerous under the serosa. The muscle of the stomach is negative for iron and shows no excess of haemofuscin, the ganglion cells of the myenteric plexus contain iron in small amount in Cases 4 and 5. The *duodenum* shows a moderate staining of Brunner's glands and numerous iron-positive phagocytes are seen in the tips of the villi but the gland cells

[1] The plates referred to in this lecture will be found between pages 344–5.

TABLE 1

Case age sex	Blood dyscrasia	Amount transfused Litres	Iron (g.)	Duration transfusions (years)	Estimated iron recovered (g.) Liver	Spleen	Case no.
M 71	Refractory megalocytic anaemia with aplastic crises	25+	12 5	8	No analyses		I
M 64	Idiopathic aplasia	27	13·5	3	5	0 4	II
F 72	Idiopathic aplasia	94	47	3	2·5	0·5	III
M 44	Osteosclerosis, aplasia	415	210	12	55	10	IV
M 67	Idiopathic aplasia	305	152	6·5	57·75	28	V

surface/volume ratio and thus encourages the cellular multi- [illegible] ount of fuscous pigment in the [illegible] the gross enlargement of the [illegible] due to three factors, (*a*) en- [illegible] general due to distension of their cytoplasm by the accumulated iron compound, (*b*) pronounced broadening of the portal tracts by massive accumulation of iron-laden macrophages, and (*c*) a degree of hyperplasia of liver cells.

Iron storage in the *pancreas* is seen in the acinus cells, in the cells lining the ducts and their glands, in the islets of Langerhans and in macrophages in the connective tissue (Plate XVIII, Fig. 4). There is no fibrosis or atrophy in four cases but in Case 3 a notable degree of fatty infiltration is present, with much atrophy of the acinous tissue. Many islets, however, have survived and some are greatly hypertrophied. Haemosiderin is stored chiefly in the basal part of the acinar cells; its distribution from lobule to lobule is irregular but it is often most abundant in acini which closely invest the islets (Plate XVIII, Fig. 4); the cells with the heaviest concentration of iron are usually somewhat shrunken and post-mortem changes add to the difficulty in distinguishing them from clusters of macrophages. The tall columnar cells lining the pancreatic ducts and the cells and ducts of the small mucous glands which open into them show a striking degree of iron staining.

In the islets of Langerhans, haemosiderin accumulates in both alpha and beta cells and occasionally is abundant. In the stroma and adjacent fatty tissue iron-storing macrophages are numerous.

The *lymph nodes* draining the liver and pancreas are so massively replaced by haemosiderin-storing macrophages that their normal structure is virtually obliterated, only small foci of cortical lymphoid tissue remaining (Plate XVIII, Fig. 5). Multinucleated giant-cells laden with iron are common; there is no discernible fibrosis.

In the cervical nodes there is considerable iron staining in the macrophages of the peripheral sinus and also in the littoral cells and macrophages of the medullary pathways.

contain much less granular iron than those of the gastric glands. In the *jejunum* only the cells lining the depths of the crypts in the upper part of the gut are slightly iron-positive. The *ileum* and *colon* are negative for iron. The smooth muscle of the gut is iron free and shows no excess of haemofuscin.

The *liver*, greatly enlarged in Cases 4 and 5, in all instances shows accumulation of haemosiderin both in the parenchymal cells and in the reticulo-endothelial cells. The flattened sinusoidal lining cells contain iron, and the Kupffer cells are even more heavily laden so that many become detached and free in the blood of the hepatic veins; others migrate from the sinusoids, some towards the central vein but mostly into the portal tracts which thus come to be very heavily infiltrated with macrophages stuffed with iron (Plate XVIII, Fig. 3). Clumps of similar cells occur around the central veins and along the walls of the interlobular veins. In these macrophages cellular details are obscured by the pigment and many of the cells may be dead but there is no evidence of iron impregnation of stromal fibres such as would occur by diffusion from dead cells. The aggregation of iron-laden phagocytes broadens the portal tracts and renders them conspicuous but there is a striking absence of sequestration of parenchyma by bands of fibrous tissue such as are characteristically found in a true cirrhosis. Pigmentation of the parenchyma reaches its greatest intensity in Case 4 where it is so abundant as to fill the cytoplasm with densely packed granules. In small foci of hyperplastic liver cells (Case 4), the granules are less densely aggregated and are seen to be orientated at the biliary pole of the cells around the canaliculi (Plate XVIII, Fig. 3). The cells lining the smaller bile ducts also contain many fine haemosiderin granules. In Case 4 the focal hyperplasia may be attributable to an attack of homologous serum jaundice one year before death but the liver shows no other evidence of damage by hepatitis. In this case the liver contained over 55 g. of elemental iron, and it must be presumed that the iron would be stored within the cells in combination, however loose, so that the weight and volume of the stored material must add considerably to the size and weight of the liver. Further, the distension of the cells by this relatively inert material alters the

surface volume ratio and thus encourages the cellular multiplication of hyperplasia. The amount of fibrous pigment in the liver is quite small. Accordingly the gross enlargement of the liver in such cases appears to be due to three factors. An enlargement of the hepatic cells in general due to distension of their cytoplasm by the accumulated iron compound. A pronounced broadening of the portal tracts by massive accumulation of iron-laden macrophages and to a degree of hyperplasia of liver cells.

Iron storage in the *pancreas* is seen in the acinar cells, in the cells lining the ducts and their glands, in the islets of Langerhans and in macrophages in the connective tissue (Plate XVIII, Fig. 4). There is no fibrosis or atrophy in four cases but in Case 3 a notable degree of fatty infiltration is present, with much atrophy of the acinous tissue. Many islets, however, have survived and some are greatly hypertrophied. Haemosiderin is stored chiefly in the basal part of the acinar cells; its distribution from lobule to lobule is irregular but it is often most abundant in acini which closely invest the islets (Plate XVIII, Fig. 4); the cells with the heaviest concentration of iron are usually somewhat shrunken and post-mortem changes add to the difficulty in distinguishing them from clusters of macrophages. The tall columnar cells lining the pancreatic ducts and the cells and ducts of the small mucous glands which open into them show a striking degree of iron staining.

In the islets of Langerhans, haemosiderin accumulates in both alpha and beta cells and occasionally is abundant. In the stroma and adjacent fatty tissue iron-storing macrophages are numerous.

The *lymph nodes* draining the liver and pancreas are so massively replaced by haemosiderin-storing macrophages that their normal structure is virtually obliterated, only small foci of cortical lymphoid tissue remaining (Plate XVIII, Fig. 5). Multinucleated giant-cells laden with iron are common; there is no discernible fibrosis.

In the cervical nodes there is considerable iron staining in the macrophages of the peripheral sinus and also in the littoral cells and macrophages of the medullary pathways.

contain much less granular iron than those of the gastric glands. In the *jejunum* only the cells lining the depths of the crypts in the upper part of the gut are slightly iron-positive. The *ileum* and *colon* are negative for iron. The smooth muscle of the gut is iron free and shows no excess of haemofuscin.

The *liver*, greatly enlarged in Cases 4 and 5, in all instances shows accumulation of haemosiderin both in the parenchymal cells and in the reticulo-endothelial cells. The flattened sinusoidal lining cells contain iron, and the Kupffer cells are even more heavily laden so that many become detached and free in the blood of the hepatic veins; others migrate from the sinusoids, some towards the central vein but mostly into the portal tracts which thus come to be very heavily infiltrated with macrophages stuffed with iron (Plate XVIII, Fig. 3). Clumps of similar cells occur around the central veins and along the walls of the interlobular veins. In these macrophages cellular details are obscured by the pigment and many of the cells may be dead but there is no evidence of iron impregnation of stromal fibres such as would occur by diffusion from dead cells. The aggregation of iron-laden phagocytes broadens the portal tracts and renders them conspicuous but there is a striking absence of sequestration of parenchyma by bands of fibrous tissue such as are characteristically found in a true cirrhosis. Pigmentation of the parenchyma reaches its greatest intensity in Case 4 where it is so abundant as to fill the cytoplasm with densely packed granules. In small foci of hyperplastic liver cells (Case 4), the granules are less densely aggregated and are seen to be orientated at the biliary pole of the cells around the canaliculi (Plate XVIII, Fig. 3). The cells lining the smaller bile ducts also contain many fine haemosiderin granules. In Case 4 the focal hyperplasia may be attributable to an attack of homologous serum jaundice one year before death but the liver shows no other evidence of damage by hepatitis. In this case the liver contained over 55 g. of elemental iron, and it must be presumed that the iron would be stored within the cells in combination, however loose, so that the weight and volume of the stored material must add considerably to the size and weight of the liver. Further, the distension of the cells by this relatively inert material alters the

in small quantity except in Case 4, where notable accumulations of iron are seen in the alveolar phagocytes and in phagocytes aggregated along blood vessels and in the stroma. Iron impregnation of the elastic tissue and reticulum of the lung parenchyma and blood vessels is absent.

Genito-urinary Tract. The *kidney* contains little iron and that mostly in phagocytes along the vascular leashes. In the renal epithelium itself iron is scanty (Plate XIX, Fig. 8), except in the ascending limb of Henle's loop and the distant convoluted tubule where the lining cells within a narrow zone may be quite heavily laden, the degree of iron storage varying from nephron to nephron. In addition, some glomeruli contain small but intense iron deposits, perhaps the result of impaction in the capillaries of iron-laden macrophages detached from the reticulo-endothelial system elsewhere.

In the *testis* there is atrophy of the seminiferous tubules and absence of spermatogenesis, iron-positive phagocytes are present in small numbers around the capillaries and arterioles in the tunica albuginea. The interstitial cells of Leydig are weakly positive, and some cells contain fine granules of iron along with the usual yellow-brown pigment.

Nervous System. The *brain and brainstem* are iron negative. In the *choroid plexuses* iron is present in large amount in the covering epithelium (Plate XIX, Fig. 9) and small aggregates of iron-rich phagocytes are present in the vascular stroma. In the *pituitary* gland iron is present in both lobes and in the pars intermedia. In the anterior lobe iron in small amount is widely distributed in the different types of epithelial cells and in the sinusoidal lining cells. The distribution of pigment in the posterior lobe is strikingly focal and may perhaps have resulted from local haemorrhage.

The *adrenal cortex* shows intense siderosis in the glomerular zone, most marked in the outermost cells, but extending into the zona fasciculata, but the reticular zone is almost devoid of haemosiderin (Plate XIX, Fig. 10). The medullary cells also contain fine iron granules irregularly distributed. The haemosiderin granules in the cortical cells are fine and closely packed, whereas those in the macrophages derived from the reticulo-endothelial

In the mesenteric lymph nodes, iron accumulation is less abundant in the reticulo-endothelial cells and massive accumulations are absent and this must be related to the distribution of iron in the small intestine.

Spleen. In all five spleens, there are focal accumulations of iron-laden macrophages in the red pulp, in the trabeculae and in the capsule, sometimes of massive extent (Plate XVIII, Fig. 6). The venous sinusoids are outlined by the haemosiderin granules in their endothelial cells but there is little or no iron impregnation of the reticulum fibres or of the elastic tissue. Collections of foamy cells are not seen in the spleen pulp in contrast to their abundance in thalassaemia.

Bone. In Case 4 the bone matrix stains positively for iron in some areas; and this reaction seems to depend on the presence of abundant iron in the adjacent marrow. Then the medullary trabeculae give a strongly positive prussian blue reaction which outlines the cement lines and sharply brings out the arrangement of the bone lamellae (Plate XIX, Fig. 7). In the cortical bone of the femur the reaction is less intense, being most distinct in the bone immediately adjacent to the Haversian canals. The varied intensity of the staining suggests that iron has been incorporated in the matrix at periods when bone formation has been active, and perhaps it is significant that iron staining of bone matrix is present only in the patient in whom absorption and new formation of bone had led to some degree of osteosclerosis. The findings in the *cellular marrow* of the different cases vary according to the nature of the dyscrasia which had rendered maintenance by transfusion necessary, but excessive iron storage is common to all. Only in Case 4 is fibrosis seen and since osteosclerosis was present in this case the fibrosis also is attributed to the primary disease and not to a secondary effect of iron excess. In Case 4 a striking feature is the presence of numerous foam cells in the marrow, not seen in the other cases. The significance of these is uncertain but their presence recalls the similar finding in thalassaemia (Whipple and Bradford, 1936).

Respiratory System. The mucous glands of the trachea and main bronchi give a faint iron reaction. The *lungs* contain iron

diabetes. The total amount of iron administered by the 94 transfusions is approximately 20 g. and to that can be added 3 g. intravenously and an unknown quantity absorbed from the gut. But the liver weighed 2,100 g. and contained 1·1 per cent of iron, i.e. his liver *alone* contained as much iron as he had received parenterally, and when we add to this the considerable quantity of iron in his other organs it is clear that his tissues contain a substantially greater quantity of iron than had been administered in the form of blood.

TABLE 2. Three cases of generalized siderosis with refractory anaemia and hypercellular marrow

All treated with iron, liver and transfusions. Hepatic iron exceeds transfused iron.

Case	History	P.M. findings	Hepatic iron
I	♂ 60; anaemia 3 years, iron by mouth and parenterally; 94 pints blood	hepatic cirrhosis; pancreas: no fibrosis	23 g.
II	♂ 44, anaemia 10 years; transfused <10 pints, iron by mouth throughout illness	hepatic cirrhosis; pancreas: no fibrosis, severe fibrosis of spleen and lymph nodes	32 g.
III	♀ 60, anaemia 15 months; hepatitis, jaundice; splenomegaly; iron by mouth and parenterally; 4 pints blood	hepatic cirrhosis; pancreas. no fibrosis	2 9 g.

Another case was that of a man of 44 with a refractory anaemia of 10 years' duration who died of cardiac failure with cirrhosis of liver and ascites. He too had a markedly *hypercellular* marrow, and generalized siderosis with iron storage in all the usual sites. The liver shows intense siderosis and a severe multilobular cirrhosis with regenerative nodules in which the liver cells contain less iron but the Kupffer cells are filled. The pancreas is laden with iron but not fibrosed. The lymph nodes showed dense fibrosis with an enormous amount of iron in macrophages and encrusting the elastic and reticular tissue and the spleen was enlarged, fibrosed and had abundant iron. These are unusual features and we have provisionally classified the

cells of the capillaries both in the cortex and medulla are large, coarse, and so densely aggregated as to obscure cellular detail. As in the liver small foci of cortical hyperplasia show less pronounced iron storage. In Case 4 amyloid substance is present in the fine stroma of the cortex and medulla.

The *thyroid gland* shows well-marked siderosis of the cells lining the vesicles, and also in the macrophages in the interstitial tissue, but stainable iron is absent from the thyroid colloid (Plate XIX, Fig. 11). Iron is more abundant in the epithelial cells lining small vesicles poor in colloid or with thin colloid and is least abundant in the cubical lining cells distended with firm ribbed colloid.

Muscle. The cardiac musculature (Plate XIX, Fig. 12) is much more severely affected than the voluntary muscles and smooth muscle was always devoid of iron. The pigment is aggregated chiefly at the poles of the nuclei, but when the deposit is heavy, the granules extend far along the fibres. The distribution of iron seems quite irregular in limb muscles but in the heart three zones can be recognized. Iron is most abundant in the subepicardial muscle, the subendocardial zone comes next, whereas the intermediate muscle contains much less than the others.

Synovial membrane from the major joints is heavily pigmented, but changes in the lining cells are absent.

Every one of these five cases of Group I was of aplastic (aregenerative) anaemia with severe marrow hypoplasia. Group II, however, consists of three cases of 'transfusional siderosis' with a very different picture (Table 2).

The first of these was one of refractory anaemia with *hyperplastic* bone marrow and over a period of about 3 years he received 94 pints of blood by transfusion. He was also treated by administration of 3 g. of saccharated oxide of iron intravenously and had received repeated and prolonged courses of iron by mouth. Not only has he pronounced iron storage in his organs, which follows the normal pattern, but there is also a well-marked cirrhosis of the liver and a minor degree of pancreatic fibrosis along with pronounced fatty infiltration. He had occasional glycosuria but blood sugar curves failed to reveal a true

diabetes. The total amount of iron administered by the 94 transfusions is approximately 20 g. and to that can be added 3 g. intravenously and an unknown quantity absorbed from the gut. But the liver weighed 2,100 g. and contained 1·1 per cent of iron, i.e. his liver *alone* contained as much iron as he had received parenterally, and when we add to this the considerable quantity of iron in his other organs it is clear that his tissues contain a substantially greater quantity of iron than had been administered in the form of blood.

TABLE 2. Three cases of generalized siderosis with refractory anaemia and hypercellular marrow
All treated with iron, liver and transfusions. Hepatic iron exceeds transfused iron.

Case	History	P.M. findings	Hepatic iron
I	♂ 60, anaemia 3 years; iron by mouth and parenterally; 94 pints blood	hepatic cirrhosis; pancreas: no fibrosis	23 g.
II	♂ 44; anaemia 10 years; transfused <10 pints; iron by mouth throughout illness	hepatic cirrhosis; pancreas no fibrosis; severe fibrosis of spleen and lymph nodes	32 g
III	♀ 60, anaemia 15 months; hepatitis, jaundice, splenomegaly, iron by mouth and parenterally; 4 pints blood	hepatic cirrhosis; pancreas: no fibrosis	2 9 g.

Another case was that of a man of 44 with a refractory anaemia of 10 years' duration who died of cardiac failure with cirrhosis of liver and ascites. He too had a markedly *hypercellular* marrow, and generalized siderosis with iron storage in all the usual sites. The liver shows intense siderosis and a severe multilobular cirrhosis with regenerative nodules in which the liver cells contain less iron but the Kupffer cells are filled. The pancreas is laden with iron but not fibrosed. The lymph nodes showed dense fibrosis with an enormous amount of iron in macrophages and encrusting the elastic and reticular tissue and the spleen was enlarged, fibrosed and had abundant iron. These are unusual features and we have provisionally classified the

case as one of refractory anaemia with hepatic cirrhosis and siderosis.

The third case in this group again exemplifies hepatic fibrosis with siderosis of liver, pancreas and other organs in a case of refractory anaemia of about 2 years' duration treated by intensive oral and parenteral iron therapy and only to a minimal extent by transfusion, amounting in all to only 4 pints of blood.

These cases are not examples of classical haemochromatosis nor can they be classified as transfusional siderosis. In all three cases the amount of iron found in the organs far exceeds that administered by transfusion. In each case the patient had been treated intensively by the administration of iron by the mouth and in two cases also parenterally. These cases resemble those recorded by Goldish and Aufderheide (1953), by Wallerstein and Robbins (1953) and by Wyatt and others; and these workers (with whom we agree) attributed the accumulation of iron to the prolonged oral iron therapy for anaemia that was refractory to treatment.

From the published cases of so-called transfusional siderosis we reached the conclusion some years ago that there were few if any cases recorded in which true hepatic fibrosis was associated with transfusional siderosis where the amount of iron stored in the liver and other organs could fairly be attributed to the transfused iron. In almost every case in which cirrhosis of liver was recorded by the author the amount of iron estimated in the organs by chemical analysis was greater than that administered in the form of blood. The excess must have been absorbed from the gut, and it is, of course, well known that the proportion of iron absorbed is greater in anaemia and especially, according to Bothwell and others (1958), when the bone marrow is hypercellular.

Recently Oliver (1959) gave an account of 22 cases receiving [illegible] stated to show [illegible]

presence or absence of hepatic fibrosis. Thus Oliver's aplastic cases 8 and 9 each received 217 pints, both showed siderosis but only one had mild cirrhosis; four cases receiving less than

100 pints showed fibrosis ranging from slight to moderate whereas the case with the largest intake, viz. 234 pints, showed only moderate hepatic fibrosis despite severe siderosis. It is difficult to be certain about the state of the marrow in all the cases recorded as showing cirrhosis and having on analysis more iron in the tissues than the transfusions would account for, but it is certainly the usual finding that these are cases with a hypercellular marrow and only rarely if ever is the marrow illustrated as acellular or aplastic.

The distribution of iron between the organs and tissues in transfusional siderosis is far from uniform and we do not know why in one case the liver will have 4 per cent of iron and the pancreas 1 per cent whereas in another the liver will have 6 per cent and the pancreas about 0·5 per cent. The increased concentration does not depend merely on a reduced size of the organ, nor is it a constant proportion of the iron administered, the total liver iron being sometimes one-quarter, in other cases almost one-half. A similar variability is seen in haemochromatosis and other iron-storage disorders. Primary hepatic carcinoma is rather more frequent in haemochromatosis than in ordinary portal cirrhosis but hepatic cancer in a case of transfusional siderosis is unknown.

In transfusional siderosis there is a notable accumulation of iron-containing phagocytes in the stroma of the portal tracts broadening them to an extent sometimes mistaken for cirrhosis. This intense accumulation of macrophages with ingested iron arises in part from continued migration of Kupffer cells but is thought by some to be in part the result of disintegration of iron-laden liver cells and thus to indicate liver damage; it is also blamed for the fibrosis by reason of its allegedly irritant action. One cannot, however, overlook the fact that cases repeatedly transfused run a considerable risk of acquiring homologous serum jaundice, residual damage from which is certainly another possible cause of cirrhosis.

HAEMOCHROMATOSIS

Iron accumulation in the portal tracts occurs also in haemochromatosis but it is of very variable degree and some cases

with abundant liver iron may show so little iron accumulation in the portal tracts that it could not be responsible for the massive degrees of fibrosis observed. It is, I think, pertinent to point out that the cause of the fibrosis of the liver in haemochromatosis is not known. Muir and Shaw Dunn (1914–15) thought that the cirrhosis of liver came first and caused some abnormality of iron metabolism which led to the great accumulation of pigment, and many have held that the two processes of cirrhosis and pigmentation of the liver were concurrent but independent. Davis and Arrowsmith (1953) have shown that the deposits of iron in haemochromatosis are not so firmly fixed in the organs as had been believed and that repeated phlebotomy can mobilize at least a substantial portion of the hepatic iron stores, the iron being used in haemoglobin synthesis. Provided care is taken to restore the plasma proteins, this has proved to be a useful method of treatment, but some cases show signs of hypochromic anaemia when the volume of blood removed is almost certainly less than the equivalent of the iron deposited in the liver, and it is likely that some of the iron becomes converted into an insoluble and unavailable form. Haemochromatosis is a disease that continues to interest physicians and pathologists on account of the obscurity that surrounds it. It is an uncommon but not exceedingly rare disease, and a problem that confronts the pathologist is to know when to apply the diagnostic label. During the past 50 years, 30 cases of fully developed haemochromatosis have been necropsied in the Western Infirmary, Glasgow, and in addition there have been 7 cases dying of other causes in which the liver has shown iron pigmentation with or without a degree of hepatic cirrhosis and with lesser degrees of pigmentation in the pancreas and other organs. Some of these cases were accepted by Muir as *formes frustes* of the disease and as evidence that hepatic and pancreatic cirrhosis preceded the pigmentation. Excessive iron storage in the liver in other forms of hepatic cirrhosis is not uncommon and the dividing line between one group and another is by no means sharp. Some emphasis has been laid by Mazur and Shorr (1948) on the occurrence of shock-like attacks in haemochromatosis which were attributed to the release of a vasodilator substance thought

to be ferritin. Amongst our 30 post-mortems there were 13 cases of sudden general abdominal pain of which 9 at necropsy proved to be due to general peritonitis for which no local focus of origin could be found in any of the usual sites such as appendix, perforation of a peptic ulcer, etc. Indeed the source of the bacterial peritonitis was not traced, and this tendency to develop peritonitis is not so well known as it should be.

Haemochromatosis is a disease of middle and later age, the amount of iron stored is large, as much as 25 g. may be present in the liver and probably as much again in the other organs, of which the pancreas, stomach, heart, thyroid, adrenal, pituitary and choroid plexus are most severely and consistently affected. The marrow contains more iron than normally, but the spleen has no excess. If we assume a normal iron content in the diet of 10–15 mg. per day, total retention would provide the necessary additional iron in less than 10 years. Petersen and Ettinger (1953) have shown by means of radio-iron studies that the absorption of iron from the gut in haemochromatosis is increased to about 20–40 per cent instead of 2–5 per cent as in normal persons, but that very little of this iron appears in the circulating red cells, and presumably most of it goes into storage. Why this happens is obscure, because as we have seen, the stored iron can be mobilized for haemoglobin synthesis when required. There is no evidence that the stored iron has ever passed through the metabolic cycle of haemoglobin and as I have pointed out the prefix *haemo*, which certainly implies a relation to haem pigments, is in fact not justified, though no doubt it is too well established to be easily discarded. If, however, haemochromatosis were merely a matter of excessive absorption of iron, then we might reasonably expect that its morbid anatomy would be identical with that of other conditions attributed to excessive absorption, e.g. siderosis in the Bantu and the siderosis resulting from prolonged iron medication by mouth. This is, however, not so and there are certain differences for which we cannot at present account (see Tables 3 and 4).

When the late A. S. Strachan, also a lecturer in Glasgow with Robert Muir, went to Johannesburg in 1923, he observed that a condition of siderosis closely resembling haemochromatosis

with abundant liver iron may show so little iron accumulation in the portal tracts that it could not be responsible for the massive degrees of fibrosis observed. It is, I think, pertinent to point out that the cause of the fibrosis of the liver in haemochromatosis is not known. Muir and Shaw Dunn (1914–15) thought that the cirrhosis of liver came first and caused some abnormality of iron metabolism which led to the great accumulation of pigment, and many have held that the two processes of cirrhosis and pigmentation of the liver were concurrent but independent. Davis and Arrowsmith (1953) have shown that the deposits of iron in haemochromatosis are not so firmly fixed in the organs as had been believed and that repeated phlebotomy can mobilize at least a substantial portion of the hepatic iron stores, the iron being used in haemoglobin synthesis. Provided care is taken to restore the plasma proteins, this has proved to be a useful method of treatment, but some cases show signs of hypochromic anaemia when the volume of blood removed is almost certainly less than the equivalent of the iron deposited in the liver, and it is likely that some of the iron becomes converted into an insoluble and unavailable form. Haemochromatosis is a disease that continues to interest physicians and pathologists on account of the obscurity that surrounds it. It is an uncommon but not exceedingly rare disease, and a problem that confronts the pathologist is to know when to apply the diagnostic label. During the past 50 years, 30 cases of fully developed haemochromatosis have been necropsied in the Western Infirmary, Glasgow, and in addition there have been 7 cases dying of other causes in which the liver has shown iron pigmentation with or without a degree of hepatic cirrhosis and with lesser degrees of pigmentation in the pancreas and other organs. Some of these cases were accepted by Muir as *formes frustes* of the disease and as evidence that hepatic and pancreatic cirrhosis preceded the pigmentation. Excessive iron storage in the liver in other forms of hepatic cirrhosis is not uncommon and the dividing line between one group and another is by no means sharp. Some emphasis has been laid by Mazur and Shorr (1948) on the occurrence of shock-like attacks in haemochromatosis which were attributed to the release of a vasodilator substance thought

Plates XVIII and XIX (*overleaf*)

Plate XVIII

Fig. 1 Case 4 Mamma Epithelium contains fine iron granules and iron-containing phagocytes are numerous in stroma (× 190)

Fig. 2 Case 4 Stomach The specialized secretory cells in the depths of the fundus glands are strongly positive for iron Iron-storing phagocytes in the lamina propria (×65)

Fig. 3 Case 4 Liver There is intense siderosis, both of the liver cells and Kupffer cells and in phagocytes within the portal tract No cirrhosis, but hyperplasia of liver cells is shown by the area of less intense iron storage (× 100)

Fig. 4 Case 4 Pancreas General architecture preserved No fibrosis Siderosis of both islets and exocrine elements (×50)

Fig. 5 Case 4 Portal lymph-node Very intense iron storage but general architecture still discernible, no gross fibrosis (×50)

Fig 6 Case 5 Spleen The Malpighian bodies are slightly pigmented in contrast to the splenic pulp, in which iron storage is abundant (×65)

Plate XIX

Fig 7 Case 4 Femur Striking iron impregnation of the bone, outlining the cement lines The marrow is aplastic with relatively little siderosis in this area (×50)

Fig. 8 Case 1 Kidney Granular iron in the epithelium of the distal part of the nephron only, the proximal loops are iron-free (×60)

Fig 9 Case 5 Choroid plexus Intense staining of the epithelium, commonly forming a 'cap' to the cell (×245)

Fig 10 Adrenal cortex Iron storage in parenchyma chiefly of zona glomerulosa, and in capillary endothelium (×50)

Fig 11 Thyroid gland The epithelium, especially of the small colloid-poor vesicles, is rich in iron (×50)

Fig 12 Case 5 Myocardium Abundant iron is present within the muscle fibres (×190)

All sections stained with potassium ferrocyanide and hydrochloric acid, and then with lissamine red

was unexpectedly common amongst the Bantu population of that area. Strachan was the first to study closely this condition and, following Mallory's work (1925) on the relation of chronic copper poisoning to haemochromatosis, Strachan was inclined to attribute the remarkable iron deposits observed in the Bantu to excessive intake of zinc, iron and other heavy metals derived from the cooking and other food utensils of the Bantu natives.

TABLE 3. Comparison of Fe content of organs (percentage dry weight)

	Normal total Fe	Haemochromatosis (Sheldon)	Bantu siderosis (Wainwright) 4 cases	Transfusional siderosis (Cappell *et al*) 4 cases
Liver	0 05–0 1	3 65	2·80	3·1
Pancreas	0 018	1 89	0 17	0 92
Spleen	0 14	0 63	2 95	2·7
Lymph node	—	7·92	4·1 (wet wt.)	4 9
Kidney	0 04	0 195	0·175	0 229
Heart	0 039	0 517	0 033	0 14
Stomach	0 045	0 226	0 02 (wet)	0 37
Duodenum	0 057	0 138	0 33 (wet)	0·358
Total iron content of organs	5 g	25–50 g	8–15 g.	50 g. +

The descriptions of Bantu siderosis given by Strachan (1929), by the Gillmans (1951), by Higginson and his co-workers (1953) and by Wainwright (1957) do not differ substantially on the morbid anatomy and histology but the authors' interpretation and their explanation of the etiology differ profoundly. All agreed that, in addition to the maximum site of pigmentation in the liver, the major sites of iron deposit are the spleen, marrow, duodenum and jejunum, all of which are relatively little pigmented in haemochromatosis, whereas the pancreas, stomach and choroid plexus, organs notably pigmented in true haemochromatosis, are virtually devoid of iron excess. These differences cannot be disregarded and they are not attributable merely to lesser degrees of iron storage, although, according to Wainwright, the total amount of iron retained in Bantu siderosis

Plate XIX

Transfusional Siderosis

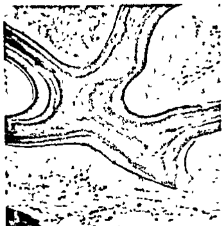


Fig. 7



Fig. 8

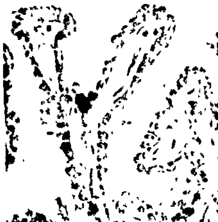


Fig. 9



Fig. 10



Fig. 11

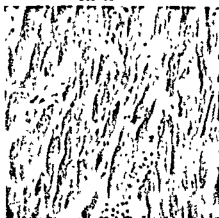


Fig. 12

Plate XVIII

Transfusional Siderosis



Fig 1



Fig 4



Fig 2

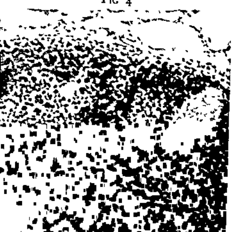


Fig 5



Fig. 3

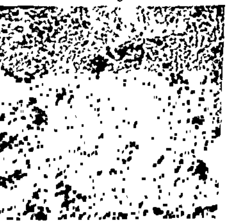


Fig. 6

is much smaller than in haemochromatosis or in our cases of transfusional siderosis, rarely exceeding 8 to 15 g. in his series.

TABLE 4 Comparison of sites of iron storage

	Transfusional siderosis	Haemo-chromatosis	Bantu siderosis	Thalassaemia
Liver	+++	+++	+++	+++
Spleen	++	(+)	+++	−
Marrow	+	+	++	(+)
Pancreas	++	++	−	++
Lymph nodes	++	++	+	++
Stomach	++	++	−	++
Duodenum	+	+	+++	(+)
Heart	+	+	(+)	+
Thyroid	++	++	+	+
Kidney	(+)	(+)	(+)	(+)
Choroid plexus	++	++	−	−

An important and possibly significant difference between idiopathic haemochromatosis and Bantu siderosis is seen in the amount of iron carried by the plasma globulin, siderophilin (Table 5).

TABLE 5. Serum iron and siderophilin levels

	Normal		Bantu siderosis		Haemo-chromatosis		Transfusional siderosis	
in μg /100 ml.	range	average	range	average	range	average	range	average
Serum iron	70–180	120	150–550	360	180–320	240	170–320	234
T.I B C.	250–380	340	390–830	590	160–340	260	250–380	280
% saturation		30		60		94+		90+

In haemochromatosis the total iron-binding capacity of the plasma is not increased but the siderophilin is constantly about 90 per cent saturated instead of the normal 30 per cent, whereas in Bantu siderosis the total iron-binding capacity is substantially elevated but the percentage saturation is only moderately increased (Gerritsen and Walker, 1953). Recently Gillman and others have confirmed this finding and have claimed that it represents a metabolic disorder which may underlie and be responsible for the increased uptake of iron from the diet found



deficiency anaemia the plasma iron is low (Fig. 13). But we are not able to say with certainty that a constant high level of plasma iron means an increased rate of transfer of iron from gut to storage depots. For example in haemochromatosis the high level of saturation of siderophilin may only reflect a greater difficulty in finding tissue receptors not already saturated and still capable of accepting iron from the plasma. The significance of the raised iron-binding capacity of the plasma in Bantu siderosis is still uncertain, but it may be a factor of significance.

Iron Storage in the Kidney

Before leaving human pathology to speak of experimental work, attention should be drawn to certain notable differences in the circumstances in which the kidneys store a large amount of iron. In the conditions of siderosis already discussed, viz. transfusional siderosis, haemochromatosis and Bantu siderosis, the kidneys contain relatively very little iron despite the enormous amounts stored elsewhere, and granular deposits seen are restricted to the distal tubules. In striking contrast, in acquired haemolytic anaemia, in paroxysmal nocturnal haemoglobinuria and all other forms of haemoglobinuria, there is abundant and *widespread deposition of iron in the proximal convoluted* tubules. In the so-called haemolytic crises of congenital spherocytosis, now recognized to be in reality episodes of failure of marrow output, and in Addisonian anaemia in relapse a similar deposit of iron occurs in the proximal tubules.

The findings in haemoglobinuric kidneys may be attributable to glomerular filtration with tubular reabsorption and degradation of haemoglobin by the proximal lining cells, but in congenital spherocytosis there is never haemoglobinuria although at times haemolysis, as judged by the persistent high reticulocyte count, may be very severe. Nevertheless the state of the kidneys so closely resembles the known results of haemoglobinuria that we may reasonably suspect in those two diseases the continued presence of a trace of haemoglobinaemia, for this [illegible]

in Bantu siderosis. They have, moreover, observed that these changes in siderophilin were present in otherwise healthy Africans in whom liver biopsy disclosed uncomplicated siderosis. There is, however, no evidence that the amount of siderophilin

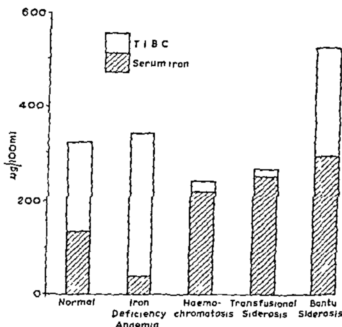


FIG. 13
Serum Iron levels and total iron-binding capacity.

or its degree of non-saturation influences the capacity to absorb iron from the gut, and Brown, Dubach and Moore (1958) state unequivocally that saturation of the plasma siderophilin does not reduce absorption of ^{59}Fe. There is no evidence that the amount of iron in the plasma is closely correlated with the rate at which iron enters the plasma from the gut or depots and leaves it in the marrow or storage depots. Normally a sudden increase of alimentary iron is followed by a rise in plasma iron levels as is seen in the estimation of plasma iron curves, and the excess is removed from the plasma within a few hours. In iron-

the plasma and then enters into the parenchymatous cells of various organs. The liver is especially active, though the suprarenals and also the kidneys take part in a small way, but with moderate doses the pancreas and stomach are not affected. With a colloidal non-diffusible substance like saccharated oxide of iron the lymph-nodes and the histiocytes of the general connective tissues are at first unaffected because the colloidal iron is separated from them by the endothelial barrier. But in the redistribution of iron that follows, iron certainly passes into the tissue fluid and lymph so that the reticulo-endothelial cells of lymph nodes and the histiocytes then contain stainable iron. No doubt this soluble form of iron is conveyed in the iron-binding β-globulin siderophilin and, since all tissue fluid contains at least a trace of protein, perhaps this is the mode of conveyance of iron to the tissue cells. It is, however, not yet known with certainty how the siderophilin comes to part with its iron to the storage cells, whether it merely dissociates and hands on ionic iron to receptors in the cells or whether the siderophilin is itself metabolized in the process, but the latter view has some powerful quantitative arguments against it (Laurell and Nyman, 1957) Despite the marked redistribution of iron throughout the body, I entirely failed to produce any lesions resembling haemochromatosis and this has since been confirmed by the more extensive work of Nissim (1953) and of Golberg and others (1957).

I therefore concluded that this substance was likely to be useful in the treatment of hypochromic anaemia by intravenous injection, but I quite failed to convince my clinical colleagues who at that time believed implicitly in massive dosage of iron by the mouth. Later, when it was found that oral iron was not without its therapeutic failures and clinical drawbacks, the

however, the disadvantage of requiring the intravenous route and thus were not so popular in general practice as a preparation suitable for intramuscular use would be. Later the iron-dextran compounds made their appearance and have been

the tubules would occur and thus the appearance of haemoglobinuria would be prevented. Fairley's (1941) observation of methaemalbumin in the plasma in Addisonian anaemia supports this suggestion and the virtual absence of plasma haptoglobins there may well be the result of long-continued depletion by minimal chronic haemoglobinaemia. In pernicious anaemia the renal threshold for haemoglobin is lowered just as it is following repeated injections of haemoglobin: this was formerly attributed to saturation of the proximal tubule cells but today chronic haptoglobin depletion, permitting free haemoglobin to reach the glomerular filtrate, would be a more likely explanation. It is of course known from experiments with tagged red cells that the erythrocytes in pernicious anaemia are short-lived, and that more rapid haemolysis plays a part in the genesis of the anaemia. But patients with aplastic anaemia often receive by transfusion blood in which some of the cells at least are equally short-lived. Yet so long as incompatibility has been avoided the picture of iron storage in the kidneys is quite different from that in haemolytic anaemia and in Addisonian anaemia in relapse. We may speculate that the intimate mechanism of red cell destruction in pernicious anaemia may be different from the normal and indeed the abundance of haemophagocytosis in the marrow and other organs supports this hypothesis.

Experimental Administration of Iron

Saccharated oxide of iron has been widely used in therapy and during the past ten years has been employed by the intravenous route for the treatment of iron-deficiency anaemia. The rationale of this practice is to be found in the experiments I conducted 30 years ago on the reactions of the tissues to this substance. I first showed that it could safely be given intravenously, that it was initially deposited in certain cells of the reticulo-endothelial system in granular form, but that almost at once two changes appear: (*a*) the heavily laden reticulo-endothelial cells begin to leave their normal situation and migrate into the blood and towards the lymphatics, and (*b*) there begins a slow conversion of the insoluble iron to a soluble form which is redistributed in

stainable iron could be found in the marrow and the therapeutic response was good (Table 7). By the same technique, however, we have shown that after the deficiency in haemoglobin has been overcome and the haemoglobin level has returned to normal, stainable iron may fail to appear in the marrow when administration is purely oral, and it would seem that although

TABLE 7. Representative examples of cases with *no* stainable iron in marrow

Sex	Age	Hb g %	Cause	Iron therapy Duration	Iron therapy Result Hb g %
M	33	5	Haematemesis	3 weeks	10·9
F	47	6·4	Microcytic anaemia MCHC 25%	3 weeks	10·1
F	54	4·8	Microcytic anaemia	13 weeks	13·7
M	49	9·5	Microcytic anaemia Koilonychia	5 weeks	14·5
M	33	5·6	Haematemesis	5 weeks	9·4
F	57	6·6	Microcytic anaemia MCHC 26%	6 weeks	14·5

enough iron may be absorbed from the gut to repair the haemoglobin deficit, the replenishment of the iron stores may be a much slower process and indeed it may virtually fail (Hutchison, Lowther and Alexander, 1954) so that the adult patient recovered from iron deficiency after haemorrhage may remain in an iron-depleted state for a long period. Accordingly there is good reason to ensure the replenishment of the iron stores in such persons by the parenteral administration of iron in one form or another, but this need not be continued beyond the point where a small deposit in the iron stores has been attained (Hutchison *et al.*, 1954).

In the experimental animal the margin between doses equivalent to the maximum therapeutic dose in man and the maximum tolerated dose is very wide. Golberg and others (1957, 1958) have pushed dosage of therapeutic iron preparations to extremes and have shown that in the rat certain changes develop, especially with iron-dextrans, resembling the effects of vitamin E deprivation. In a later series of experiments Golberg and Smith (1959) have shown that severe iron overload of the rat liver induces an

widely used both for experimental and for therapeutic purposes (Cappell *et al.*, 1954). These are rapidly absorbed from the site of injection along the lymphatics and are quickly used in the building of new haemoglobin. Even in the largest doses at a therapeutic level they appear to be virtually harmless, at least so far as the production of lesions resembling haemochromatosis is concerned.

TABLE 6. Representative cases with abundant iron in marrow

Sex	Age	Hb. g.%	Cause	Iron therapy
F	69	11 3	uncertain	900 mg i.v. no improvement
M	58	8 2	chronic nephritis	oral and i.v. iron no improvement
M	60	6·3	refractory macrocytic anaemia	oral and i.v. iron no improvement

Nevertheless, because of doubts about the possible fibrogenic action of iron absorbed from the gut or administered parenterally, it became highly desirable, as soon as an iron preparation suitable for intramuscular use was available, to try to ensure that patients would not be given large quantities of iron *parenterally* unless they were in need of therapeutic iron. Accordingly we tried to find some criterion to assess the patient's real need for iron, more accurately than merely by the recognition of hypochromic anaemia. A careful correlation of the therapeutic response to iron with the presence or absence of histologically demonstrable iron in the initial marrow aspirates (Hutchison, 1953) led to the conclusion that the administration of iron by any route was unnecessary if the bone marrow were shown to contain any considerable quantity of iron histologically demonstrable by the prussian blue method. In Addisonian anaemia, in haemolytic and hypoplastic anaemias, the marrow was almost invariably found to contain abundant iron and the therapeutic response to iron was minimal (Table 6), whereas in the hypochromic anaemia of iron deficiency not a granule of

sequence of hepatic fibrosis and its degree would depend on secondary factors such as the composition of the diet.

Alimentary Absorption of Iron

Since McCance and Widdowson (1937) showed that there is but little real excretion of iron once incorporated into the tissues, it is clear that there must normally be an effective control to prevent excessive absorption from the average diet containing about 15 mg. of iron as compared with a daily loss of about ½–1½ mg. in men (Moore and Dubach, 1956) and somewhat more in women in the reproductive period. Absorption of food iron is of course of fundamental importance because it is the source of all the iron normally absorbed and also of the abnormal deposits found in some of the conditions I have described. Absorption of about 10 per cent of the food iron takes place in the stomach and duodenum, the iron being in the ferrous state, but absorption is considerably improved, up to about 20 per cent, by the simultaneous administration of large doses of ascorbic acid which is presumed to act by maintaining the iron in the essential reduced state. For a time it was believed that the iron of haemoglobin compounds was not absorbed because they were not sufficiently broken down during digestion. Black and Powell (1942) and Callander and others (1957) have, however, shown conclusively that the iron of haemoglobin is absorbed to about the same amount as the average of the dietary iron, i.e. 10 per cent in normals and 20 per cent in iron-deficient subjects. Indeed Johnston and others (1948) claimed that the addition of beef to a standard diet almost doubled the absorption of iron therefrom; whereas Josephs (1958) has stated that even spinach, a vegetable especially rich in iron, may reduce the absorption of alimentary iron by reason of a chelating action in the gut. On the other hand ferric sodium versenate has been found to be absorbed as well as ferrous sulphate in patients with hypochromic anaemia (Will and Vilter, 1954).

The mechanism of the increased uptake of iron in anaemic states is uncertain. It is well attested in anaemia from acute blood loss, when it appears only some days after the bleeding,

enhanced susceptibility to the harmful effects of various dietary imbalances. Deprivation of protein and of vitamin E and the addition of ethionine to the diet of rats all brought about striking pathological changes with more rapid deposit of parenchymal iron and accelerated migration of Kupffer cells to the portal tracts; portal fibrosis rapidly developed.

In contrast to transfusional siderosis ceroid pigments appeared in the organs in large amount and were widely distributed. Hepatic cirrhosis, often of severe degree, developed under such conditions especially if two adverse factors were combined. Golberg and Smith conclude that the liver overloaded with excessive quantities of iron exhibits a diminished resistance to the action of toxic agents and deficient diets, and they have shown that heavy iron storage is accompanied in rats and mice by quantitative alterations in certain hepatic enzyme systems. These careful experimental observations seem to me to offer the hope of analysing ultimately the relationship of iron storage and hepatic damage in man, especially in the Bantu. In relation to transfusional siderosis one can only speculate whether the simultaneous administration of much plasma protein may be a good thing for the patient rather than the administration of packed red cells, and clearly it would be wise in all such cases not to give iron by the mouth and to pay special attention to the diet, although the importance of tocopherols in human nutrition is less clearly defined. In the naturally occurring human sideroses Golberg and Smith consider that abnormal and excessive iron-binding in the liver may be the initial step, arising from disordered metabolism of the liver cells and they suggest that excessive formation of a hypothetical iron-binding material may be concerned. A similar view has been expressed [illegible] the Bantu that dietary [illegible] ver [illegible] ue to continuing malnutrition in adult life. Perhaps both views contain some truth and the recent experimental work seems to bear this out, as Golberg and Smith suggest. In their view the increased uptake of iron from the intestine could be a direct consequence of the increased fixation of iron in the liver and the

mucosal cells and transported in the plasma to the sites of utilization or storage. The mucosal block theory of the control of iron absorption has, however, been criticized as unphysiological and Brown, Dubach and Moore (1958) emphasize that the conditions required for the demonstration of an effective mucosal block are highly artificial and operate only when much larger quantities of iron are present in the gut than are likely to occur in any normal food intake. It seems clearly proven that neither the total amount of siderophilin nor the degree of saturation of the iron-binding capacity influence directly the uptake of iron from the gut, and we see in haemochromatosis a normal or low level of siderophilin, almost fully saturated, associated with increased uptake and increased daily turnover of iron from a normal daily intake in the food, whereas in Bantu siderosis we see a very high siderophilin level only one-third saturated despite a dietary intake of as much as 200 mg. per day. The problem of the control of iron absorption is far from solved and much remains to be discovered. There are still plenty of research problems awaiting solution.

Conclusion

In this brief review of some of the unsolved problems that confront the pathologist I have been able to touch on only a limited number of aspects but I hope that I have perhaps indicated that the contribution of morbid anatomy to the advancement of knowledge is not yet finished. In recent years biochemistry has made enormous strides and has made clear many obscurities in clinical and pathological problems. The next stage will, I hope, be to transfer to intact cells the facts recently elucidated on the structure and function of its individual parts, for example, by the use of cell homogenates. Let us look forward then to the day when histochemistry will enable us to examine precisely the processes going on in the cell organelles and by more detailed analysis to detect the reasons underlying the differences I have pointed out.

and has therefore been attributed to depletion of the iron stores (Hahn *et al.*, 1943) rather than to the anoxia. Yet in Addisonian anaemia with abundant iron stores Dubach and others (1948) found that iron was well absorbed but was not incorporated into haemoglobin as quickly as usual.

Bothwell, Pirzio-Biroli and others (1958) have stated that normally the amount of iron absorbed is inversely proportional to the amount of iron already stored and directly proportional to the activity of haemopoiesis, and Gillman and others (1958) have claimed that periodic anoxia from reduced atmospheric pressure leads to a noteworthy increase in iron uptake.

When red cell formation is artificially stimulated by almost any means, e.g. by bleeding, by haemolytic agents or even by administration of cobaltous chloride, the plasma acquires the property of stimulating erythropoiesis in normal animals. This is attributed to an extractable substance, erythropoietin, but it does not appear to have been determined as yet whether erythropoietin directly influences the alimentary uptake of iron.

In normal circumstances, however, there must be some mechanism to regulate the amount of iron absorbed. We have seen that the greatly increased iron content of the organs in haemochromatosis could be attained in about 10 years if as little as 10 mg. per day were absorbed and retained, and this quantity is well within the capacity of the diet to provide. Indeed in haemochromatosis we must accept, without being able to explain, the daily absorption of the increased amount from a diet presumably normal in iron content and we fall back on the explanation of an inborn error of metabolism. But normally man must have some mechanism to prevent this from happening and the nature of this is still obscure.

Hahn and others (1939, 1943) elaborated McCance and Widdowson's suggestion of intestinal mucosal control, and later Granick (1949, 1955) expanded this into his well-known mucosal block theory. In this it is postulated that the intestinal mucosa elaborates an intracellular receptor, apoferritin, which is quickly saturated by orally administered iron by conversion to ferritin so that no further uptake of iron from the lumen takes place until some of the iron had been removed from the

JOHNSTON, F. A., FRENCHMAN, R. and BOROUGHS, E. D. (1948). *J. Nutr.* 35, 453.
JOSEPHS, H. W. (1958). *Blood*, 13, 1.
LAURELL, C. B. and LYMAN, N. (1957). *Blood*, 12, 493.
McCANCE, R. A. and WIDDOWSON, E. M. (1937). *Lancet*, 2, 680.
MALLORY, F. B (1925). *Amer. J. Path.* 1, 117.
MALLORY, F. B., PARKER, F. and NYE, R. N. (1921). *J Med. Res.* 42, 461.
MAZUR, A and SHORR, E. (1948). *J. Biol. Chem.* 176, 771.
MOORE, C. V. and DUBACH, R. (1956). *J. Amer. med. Ass.* 162, 197.
MUIR, R. (1909). *Studies on Immunity*, London.
MUIR, R. and McNEE, J. W. (1912). *J. Path. Bact.* 16, 410.
MUIR, R. and SHAW DUNN, J. (1914/15). *J. Path. Bact.*, 19, 226.
MUIR, R. and SHAW DUNN, J. (1915). *J. Path Bact.* 19, 417.
MUIR, R. and SHAW DUNN, J. (1915). *J. Path Bact.* 20, 41.
NISSIM, J. A. (1947). *Lancet*, 2, 49.
NISSIM, J A. (1953). *J. Path. Bact.* 66, 185.
OLIVER, R. M. (1959). *J. Path. Bact.* 77, 171.
PETERSEN, R. F. and ETTINGER, R. H. (1953). *Amer. J Med* 18, 518.
PIRZIO-BIROLI, G., BOTHWELL, T. H. and FINCH, C. A. (1958). *J. Lab Clin. Med* 51, 37
SHELDON, J. H. (1935). *Haemochromatosis*, London.
STRACHAN, A. S. (1929). M.D. thesis, University of Glasgow.
J. Med. 14, 256.
ial. 9, 279.
WILL, J. J. and VILTER, R. W. (1954). *J. Lab. Clin Med* 44, 449
WYATT, J P. (1956). *Arch. Path.* 61, 42, 56.
WYATT, J. P MIGHTON, H. K. and MORAGUES, V. (1950). *Amer J Path.* 26, 883.
ZELTMACHER, K. and BEVANS, MARGARET (1945). *Arch intern Med* 75, 395.

REFERENCES

BLACK, D. A. K. and POWELL, J. F. (1942). *Biochem. J.* 36, 110.
BOTHWELL, T. H., PIRZIO-BIROLI, G. and FINCH, C. A. (1958). *J. Lab. Clin Med.* 51, 24.
BROWN, E. B., DUBACH, R. and MOORE, C. V. (1958). *J Lab. Clin. Med.* 52, 335.
CALLANDER, S. T., MALLETT, B. J. and SMITH, M. D (1957). *Brit. J. Haemat.* 3, 186.
CAPPELL, D. F. (1929). *J. Path. Bact.* 32, 595.
CAPPELL, D. F. (1930) *J Path. Bact.* 33, 175.
CAPPELL, D F., HUTCHISON, H E, HENDRY, E. B. and CONWAY, H. (1954). *Brit med. J.* ii, 1255.
CAPPELL, D. F., HUTCHISON, H. E. and JOWETT, M. (1957). *J. Path. Bact.* 74, no. 2, 245.
DAVIS, W. D. and ARROWSMITH, W R. (1952). *J Lab. Clin. Med.* 39, 526.
DAVIS, W. D. and ARROWSMITH, W. R. (1953). *Ann. intern Med.* 39, 723.
DUBACH, R, CALLANDER, S. T. and MOORE, C. V. (1948a). *Blood,* 3, 526.
DUBACH, R., CALLANDER, S. T. and MOORE, C. V. (1948b). *Blood,* 3, 1201.
FAIRLEY, N. H. (1941). *Quart J. Med* 10, 95.
FINCH, S. C. and FINCH, C. A. (1955). *Medicine,* 34, 381.
GERRITSEN, TH. and WALKER, A R. P. (1953). *S. Afr. med. J* 27, 577.
GILLMAN, J. and GILLMAN, T. (1951). *Perspectives in Human Malnutrition,* New York
GILLMAN, TH., CANHAM, P. A. S. and HATHORN M. (1958). *Lancet,* 2, 557.
GILLMAN, TH., LAMONT, N. M. and HATHORN, M. (1957). *Lancet,* 2, 173.
GOLBERG, L. and SMITH, J. P. (1958). *Br J. exp Path.* 39, 59
GOLBERG, L and SMITH, J. P. (1959). *Tr. Path Soc Gt Br* January.
GOLBERG, L., SMITH, J. P. and MARTIN, L. E. (1957). *Br. J. exp. Path* 38, 297.
[illegible]
[illegible] (1939). *J. exp Med* 69, 739.
HAHN, P. F., BALE, W F., ROSS, J. F., BALFOUR, W. M. and WHIPPLE, G. H. (1943) *J. exp Med* 78, 169.
HIGGINSON, J, GERRITSEN, TH and WALKER, A R. P (1953). *Amer. J Path.* 29, 779
HIYEDA, K. (1939). *Jap. J med. Sci* 4, 91.
HUTCHISON, H E. (1953). *Blood,* 8, 236.
HUTCHISON, H. E. (1954) *Glasg med J.* 35, 217
HUTCHISON, H E, LOWTHER, C. P and ALEXANDER, W D (1954) *J Clin. Path.* 7, 281.

Before the year 1957 a few cases had been reported, in whom the association of nephrocalcinosis, urinary calculi and oxaluria had occurred, and it is probable that the first was described by Lepoutre (1925), and again by Laas (1941), though these descriptions only record the renal findings. Further cases were reported by Wohl (1942), Vischer (1947), Vaughan, Sosman and Kinney (1947), Davis, Klinberg and Stowell (1950), Ostry (1951), Mulloy and Knutti (1951), Zollinger and Rosenmund (1952) and Chou and Donohue (1952). The identity of some of these cases may be in doubt for it was not until 1953 that Newns and Black (1953) demonstrated the association by chemical measurement of high urinary oxalate with calcium oxalate nephrocalcinosis in life, and in 1954 that Aponte and Fetter reported twins, who died with oxalosis and in which hyperoxaluria had been previously demonstrated. Post-mortem reports of examples of oxalosis have been reported since that time by Burke, Baggenstoss, Owen, Power and Lohr (1955), by Neustein, Stevenson and Krainer (1955), by Dunn (1955), by Lund and Reske-Nielsen (1956) and by Edwards (1957). There is no evidence available in these reports, however, of any known association with previous hyperoxaluria, though it seems highly probable that it was present.

In 1957, Archer, Dormer, Scowen and Watts (1957a) described two cases of hyperoxaluria with urinary calculi and nephrocalcinosis in which repeated observations over more than two years had shown a constant and excessive excretion of oxalate Further observations on these patients were published in 1958a. Both these patients maintained a high excretion until their death and at post-mortem showed the typical features of generalized oxalosis (Scowen, Stansfeld and Watts, 1959).

Myers (1957) reported two families of familial oxaluria. One of these was the family reported by Newns and Black earlier. The second family we were allowed to investigate further. The twin girls and their brother, all of whom had renal calculi, showed a normal output of oxalate and therefore cannot be examples of hyperoxaluria.

Öigaard and Soderhjelm (1957) reported a family of six children of whom three of the boys and one girl had developed

XX

Hyperoxaluria

E. F. SCOWEN

THE excretion of excessive amounts of oxalate in the urine may occur in several conditions. The increase, however, is relatively small except in one condition, which has been variously called primary oxaluria (Archer, Dormer, Scowen and Watts, 1957a), idiopathic oxalate nephrocalcinosis (Aponte and Fetter, 1954), endogen bedingte calcium-oxalatniere (Wohl, 1942). Primary hyperoxaluria is preferable, since it refers to the constant essential feature of the disease, a continuous high excretion of oxalate in the urine, which is present from birth and may give rise to several clinical and morbid anatomical appearances. It, also, serves to differentiate the abnormal excretion of oxalate, which can occur in other conditions, especially the transient rises produced by excessive ingestion of oxalate or its precursors, which may be termed secondary hyperoxaluria.

The normal range of urinary oxalate excretion varies between 20–40 mgm. per 24 hours (Archer, Dormer, Scowen and Watts, 1957b). In primary hyperoxaluria the increase is persistent. It fluctuates from day to day but exceeds 100 mgm. per 24 hours. In secondary hyperoxaluria the increase is considerably less than in the primary condition and when secondary to exogenous dietetic causes is ephemeral.

The term oxalosis should be used to designate a generalized deposition of oxalate in the body and though it occurs in many instances of primary hyperoxaluria, perhaps even as a terminal entity, it may well be that this morbid anatomical appearance can occur in other conditions.

Hall). There can no longer be any doubt that the disease is an inherited abnormality.

From the earlier case reports there appeared to be an excessive proportion of males, but the increase in the numbers now available for study has altered this picture, and there is now no significant difference between male and female incidence.

not cause any signs or symptoms directly, but clinical manifestations become apparent once calculus formation has occurred (Plate XX, Fig. 1).[1]

Urinary calculi occur early and are usually present within one to four years of age. Later incidence can, however, occur and has been reported occurring as late as 31 years of age, though first calculus had occurred at the age of 8 years (Hodgkinson, 1958). *Once calculus formation has occurred the clinical* course appears relentless, though liable to considerable variation in the rate of progression. In defiance of treatment, either surgical or preventive, the quiescent intervals are punctuated more and more frequently by episodes of colic or haematuria from recurrent calculi and complicated by frequent or infrequent attacks of urinary infection. Nephrocalcinosis supervenes sooner or later and renal function fails. Some survive through years of chronic failure with polyuria and chronic disability. Others succumb with rapidly advancing uraemia, the final episode mimicking the acute onset in babies. A few survive much longer to be the victims of advancing hypertension, papilloedema, renal and often heart failure.

The outlook with calculus formation in infancy is gloomy indeed, though some survive for twenty years or more. This course, unfortunately, appears to be the most usual behaviour of the disease. The later incidence of calculus formation in adolescence is not at present explained, for the oxalate excretion does not seem to be materially different from the other group, nor does the prognosis. Calculus formation has undoubtedly occurred in later life with a continuous and high oxalate output.

[1] The plates referred to in this lecture will be found between pages 376–7.

nephrocalcinosis. Three had died of renal failure and one, though still alive, had evidence of gross renal impairment. The post-mortem showed gross oxalate deposits in the kidneys and in one instance oxalate deposits in the thyroid. It is unfortunate that the rest of the organs were not examined and no estimations of urinary oxalate reported, but it seems extremely probable that they represent true examples of the condition and the high familial incidence is of great interest.

Further accounts of post-mortem oxalosis have been added by Hollósi (1957), Neiman, Rauber, Pierson and Gentin (1957), Simkó (1957), Katsuni and Sandbank (1959), and Scowen, Stansfeld and Watts (1959).

The Clinical Manifestation of Primary Hyperoxaluria

The rarity of the disease has previously prevented the establishment of a precise clinical picture. In 1957 we were able to find 32 cases reported, including two which had been extensively investigated by us (Archer, Dormer, Scowen and Watts, 1958a). Most of the reports were purely of morbid anatomy and the clinical information was scanty or lacking. In most of them evidence of the aetiology was not forthcoming, although the post-mortem findings were strongly suggestive of the diagnosis. The available data did allow a reasonable probability that ten of these represented true examples of primary hyperoxaluria but only five of these had been proven by demonstrating the presence during life of persistently high oxalate excretion in the urine.

The incidence of another suggestive example though unproven in a sibling in the case reported by Newns and Black (1953), the occurrence in twins reported by Aponte and Fetter (1954) and another sibling in this family dying from a similar condition, strongly suggested a familial character to the disease.

Archer, Dormer, Scowen and Watts (1958b) reported genetic studies in three families. This study has now been extended to eight families, in whom we have been able to study fifteen examples of this condition, five of which occurred in one family, eleven children having survived infancy (Scowen, Watts and

when present are seen only in the kidney. The kidney may appear smaller than usual, there may be calculus, but more frequently a diffuse deposition of crystalline material, which on analysis proves to be calcium oxalate monohydrate.

Histologically the changes are widespread and consist of diffuse depositions of similar calcium oxalate deposits in many organs, without evidence of reaction. The deposits are most intensive in the kidney and in this acute and infantile form appear largely on the tubules, which are dilated, with flattened epithelial lining.

Crystalline deposits are found extensively elsewhere, particularly in the myocardium, in the bone and in cartilage. They may also occur in the brain, in thymus, in lung and in the thyroid gland. In the male deposits may be found in the testis. Some deposit may, also, occur in the walls of arteries, though this occurs more floridly, later in life. The easiest method of detecting these crystals is by the use of the polarizing microscope, as they exhibit a marked birefringence. The crystals do not stain with haematoxylin, but show an intense black colour with the von Kossa technique (Carson, 1951; Chou and Donohue, 1952; Burke *et al.*, 1955; Dunn, 1955; Hollósi, 1957).

When death occurs either later in life from the disease or when renal complications have been present for some years before death the appearances may be modified.

In some circumstances the results of hypertension and arterial degeneration may be gross and death may have occurred from pulmonary oedema. In others this is lacking, but there may be evidence of renal osteodystrophy and parathyroid hyperplasia. These changes do not hide the essential pathology, indeed it may enhance them, but it must be emphasized that the oxalate deposits other than in the kidney may not be observable to the naked eye.

From examination of the reported literature and our own pathological material, it is obvious that the main pathological lesions are always found in the kidney. The kidneys are reduced in size and often intensely scarred. The capsule is adherent and thickened. On section a gritty feeling is apparent. The cortex and medulla both show reduction in size. The cortex shows

The instances are few, the ultimate prognosis at present unknown and the identity with primary hyperoxaluria undetermined.

From the familial studies (Scowen, Watts and Hall, 1959) it has become apparent that siblings may show a high oxalate output, equal to affected members, and yet show no clinical or radiological abnormality. The oldest of such is at present only in the early teens. Whether or not the hyperoxaluria can continue indefinitely without further manifestations remains to be seen.

There remains one other group: the baby who presents at but a few months of age acutely ill with convulsions and vomiting, severe acidosis, often oliguria with progressive renal failure and death in a few days to a few weeks. At post-mortem the classical findings of oxalosis are demonstrable, with extensive calcium oxalate nephrocalcinosis. It would seem probable, if *these children are primarily hyperoxaluric, that initial dehydra*tion and sodium depletion might well impair their capacity to excrete oxalate. The mobilization of calcium and hypercalcuria would result in gross and rapid renal obstruction with insoluble oxalate. A vicious circle thus established would produce severe progressive renal failure with oxalate retention.

It is unfortunate that no reports of these cases give any information about the urinary oxalate content during life, for the pathological features strongly suggest its identity with primary hyperoxaluria. Proof is still awaited. The common clinical features can be confused with the many conditions, which produce nephrolithiasis and nephrocalcinosis. The only way in which the diagnosis can be firmly established is the demonstration of the gross hyperoxaluria by quantitative analysis.

The Pathological Appearances of Oxaluria

The morbid anatomical features not unexpectedly differ with the age of death and presumably with the duration of the disease and the resulting changes secondary to prolonged renal damage. In young children and in some young adults who have died accidentally the gross marked anatomical changes are few, and

The testis shows crystals scattered throughout the seminiferous tubules, but much larger aggregates are seen in the ducts of the rete testis (Plate XXVI, Fig. 10). These aggregates are often mingled with fibroblasts and are sometimes overgrown with epithelium. The bony deposits seem very variable, there are sometimes small deposits only, chiefly on the surface of cancellous bone trabeculae (Plate XXVI, Fig. 11), but at times the deposits may be much more extensive, in particular in the bone and cartilage at the growing epiphyseal plates. Small deposits may occur in many other organs, but when present appear insignificant with the major deposits elsewhere. The X-ray diffraction of the crystalline deposits has been compared with an authentic specimen of calcium oxalate monohydrate and shows identical features (Plate XXVII, Fig. 12) (Scowen *et al.*, 1959).

It is of interest that the cerebrospinal fluid may show a gross increase in the oxalate content. It is the only condition known at present in which this occurs (Scowen *et al.*, 1959). There may also be a high oxalate content in effusions in serous cavities at post-mortem.

The available evidence suggests that this florid deposition of oxalate occurs rapidly and may be associated with the terminal renal blockage, for in one case tissue removed from the kidney 6 years before death showed no oxalate crystals although recurrent calculi had occurred over 20 years. Six years later at post-mortem after a short terminal illness the classical changes of oxalosis were demonstrated (Scowen, Stansfeld and Watts, 1959). In addition the old scarred areas of the kidney were the only areas free from oxalate deposits. Finally in the acute juvenile forms of oxaluria, the intense deposition of oxalate in the kidney shows no evidence of chronic renal destruction and presumably occurs as an acute episode and it seems probable that the diffuse deposits elsewhere must, also, develop acutely.

The peculiar microscopic lesions in the skeletal muscles show histologically as a focal granular degeneration of muscle fibres, unaccompanied by inflammatory reaction. The nature of this change remains obscure, and further investigation of this change and other muscular changes is required.

small whitish flecks and radiating streaks of crystalline material are seen in the medullary pyramids. There are frequently multiple calculi of all sizes present and these may extend from the pelvis into the ureter. Obstruction may have led to dilatation of the pelvis or ureter, though this seems to be unusual. No other gross marked change is usually seen of the oxaluria, though many secondary changes may be observable, if complications have ensued. Attention, however, has been drawn to the voluntary muscles (Scowen, Stansfeld and Watts, 1959), for some of these, especially the quadratus lumborum, have shown a curious cross-banding of the muscle fasiculi with regular pale streaks at regular intervals a few millimetres apart.

Histologically the lesions are much more widespread and consist as in the juvenile cases of widespread deposition of crystalline material. The crystals may be aggregated as a rosette with radial striation or fissuring, but may be in wedge shapes, or as blunt needles in small sheaves or singly. The deposition in the kidney is largely tubular, it occurs in both proximal and distal tubules, though chiefly proximal. Large aggregates are found in the ducts of Bellini near the tips of the pyramids. Occasional clusters are found in glomeruli and in scarred areas crystals lie interstitially. These were presumably originally intratubular (Plates XXI, XXII, Figs. 2 and 3). There may be, also, the changes resulting from previous recurrent pyelonephritis and hypertension. Crystals are found extensively in the myocardium (Plate XXIII, Figs. 4 and 5). Some of the larger aggregates appear interstitial, but the smaller areas lie over separate fibres and may be intracellular.

There is extensive deposition in the arterial system, with a marked predilection for vessels having a well-developed muscular media. Apart from occasional and isolated crystals little deposit occurs in vessels, the media of which is largely elastic tissue (Plate XXIV, Fig. 6). The large muscular vessels show a ring of crystals in the deeper layers of the tunica media (Plate XXV, Figs. 7 and 8). The smaller vessels show isolated deposits and do not form the ring seen in the larger vessels (Plate XXVI, Fig. 9). This arteriolar change occurs diffusely throughout the body.

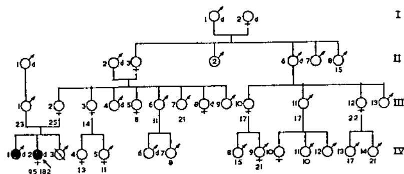


Fig. 13. Pedigree of family in which the sib of a clinically proven case of primary hyperoxaluria died from renal failure associated with recurrent urinary calculi. The figures underneath the symbols record the 24-hour oxalate output in mgm per cent

7 years. A study of the clinical records and history make it highly probable that this was an example of primary hyperoxaluria. One other sib is alive and well at 13 years of age and skiagrams do not show any renal opacities, yet the oxalate excretion is of the same order as the grossly affected members of the family. Investigation of the family (Fig. 14) has not revealed

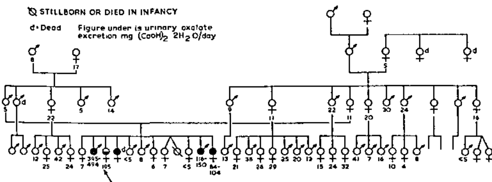


Fig 14 Pedigree of family in which multiple chemically proven cases of primary hyperoxaluria have arisen within a sibship. The figures under the symbols refer to the daily oxalate output in mgm. per 24 hours.

Observations on the Genetic Basis of Primary Hyperoxaluria

Although previous authors had used such terms as familial idiopathic oxalate nephrocalcinosis (Aponte and Fetter, 1954) and familial oxaluria (Myers, 1957) to describe the condition now called primary oxaluria, no family studies had been undertaken. The evidence for the familial character depended on the incidence in twins on one occasion with the possibility of one other sibling involvement. The additional family reported by Myers has already received comment and does not belong to this group.

Gram (1932) published an extensive pedigree of a familial incidence of oxalate calculi. No oxalate determinations were reported and the clinical course as far as it can be judged does not suggest that the cause was a primary hyperoxaluria. A previously unreported family which was grouped with the case of Newns and Black (1953) by Myers has been investigated by us. The three affected sibs, their brother, both parents and a maternal uncle showed no abnormality in oxalate excretion. [illegible] and Black had reported two affected sibs, but in one only was the urinary oxalate measured and found to be high.

It was not until 1958 that Archer, Dormer, Scowen and Watts (1958b) reported a full family study of three families. In two of these only one member was affected by the disease and no abnormally high level of urinary oxalate was discovered among the sibs, parents, first cousins, uncles, aunts and nine distant relatives. In the third family (Fig. 13) a history strongly suggestive of primary hyperoxaluria was obtained in the case of one sib of the propositus, otherwise the findings did not differ from the other two families. There was no history of consanguinity in any of the three families.

Since that time we have had the opportunity to study five more families (Scowen, Watts and Hall). One of these provides abundant and clear-cut proof of the affection of multiple sibs; for of ten sibs of the propositus, two are alive, but have renal calculi and both show persistent and high oxalate excretion in the urine. One died with bilateral renal calculi at the age of

any affected members and there is no suggestion of parental consanguinity. One other family is of great interest for the parents are first cousins. In this family one sib out of four is affected (Fig. 15).

The remaining three families gave similar pictures to the first three investigated. In one two sibs are affected. There are no unaffected sibs. The rest of the family show no abnormality and there is no parental consanguinity. The remaining two families show only the propositus affected.

The evidence for the occurrence of primary hyperoxaluria among sibs is conclusive. The presence of consanguinity in the parents of one case in a small series is a highly suggestive feature. These facts together strongly suggest that the disease is produced by the operation of a rare recessive character.

All the previous information available would fit with this contention, except for the brief report of a family by Shephard, Krebs and Lee (1958). In this family the mother and maternal grandfather of the propositus, neither of whom shows clinical evidence of renal disease, excrete 105–85 mgm. oxalate and 61 mgm. oxalate respectively per 24 hours. Shephard (1958b) gives a top normal for his method of 55 mgm. If further investigation confirms these findings there would appear to be a different mode of inheritance from those we have studied. It would appear, also, in this form to produce a much more benign clinical course. Previous mention has been made of one girl, aged 13 years, who at present appears unaffected, yet the excretion of oxalate in her urine is quantitatively similar to her affected sibs.

The survey of sex incidence of our eight families does not show any sex predilection, although previous reports would indicate a predominance of affected males.

Studies in Urinary Excretion of Oxalate

Until the present investigations were undertaken little information was available about the oxalate content of normal urine. The previous reports contained information only on isolated specimens of urine (Barrett, 1942, Powers and Levatin, 1944; Lamden and Chrystowski, 1954). There was no evidence of

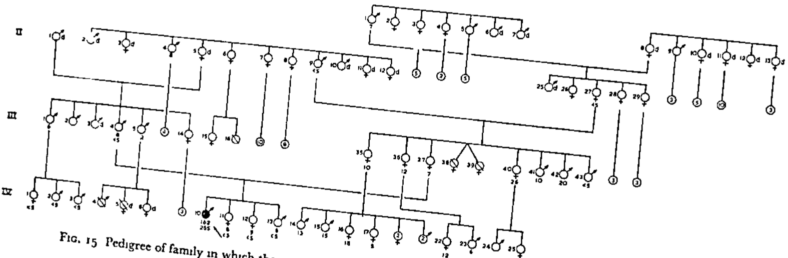


FIG. 15 Pedigree of family in which the parents of the propositus were first cousins. The figures under the symbols refer to the urinary oxalate output in mgm per 24 hours.

any affected members and there is no suggestion of parental consanguinity. One other family is of great interest for the parents are first cousins. In this family one sib out of four is affected (Fig. 15).

The remaining three families gave similar pictures to the first three investigated. In one two sibs are affected. There are no unaffected sibs. The rest of the family show no abnormality and there is no parental consanguinity. The remaining two families show only the propositus affected.

The evidence for the occurrence of primary hyperoxaluria among sibs is conclusive. The presence of consanguinity in the parents of one case in a small series is a highly suggestive feature. These facts together strongly suggest that the disease is produced by the operation of a rare recessive character.

All the previous information available would fit with this contention, except for the brief report of a family by Shephard, Krebs and Lee (1958). In this family the mother and maternal grandfather of the propositus, neither of whom shows clinical evidence of renal disease, excrete 105–85 mgm. oxalate and 61 mgm oxalate respectively per 24 hours. Shephard (1958b) gives a top normal for his method of 55 mgm. If further investigation confirms these findings there would appear to be a different mode of inheritance from those we have studied. It would appear, also, in this form to produce a much more benign clinical course. Previous mention has been made of one girl, aged 13 years, who at present appears unaffected, yet the excretion of oxalate in her urine is quantitatively similar to her affected sibs.

The survey of sex incidence of our eight families does not show any sex predilection, although previous reports would indicate a predominance of affected males.

Studies in Urinary Excretion of Oxalate

Until the present investigations were undertaken little information was available about the oxalate content of normal urine. The previous reports contained information only on isolated specimens of urine (Barrett, 1942, Powers and Levatin, 1944; Lamden and Chrystowski, 1954). There was no evidence of

possible daily fluctuation and no knowledge of changes in *excretion which might occur from changes in diet or dietary* intake of oxalate.

The methods available for determining the oxalate content of biological fluids were difficult to apply to a large number of simultaneous determinations. A modification to simplify the standard analytical procedure fortunately proved satisfactory and was used for all the subsequent work (Archer, Dormer, Scowen and Watts, 1957b). This method allows a recovery from the urine of over 90 per cent of oxalate, when the concentration does not differ materially from the normal range. When the oxalate content is pathologically high the proportion of recovery is materially higher than 90 per cent.

Normal healthy male volunteers aged 19–21 years, and weighing from 72 to 79 kg., were used to examine the normal oxalate excretions. They were given a normal mixed, but repetitive diet, the oxalate and calcium content of which was determined directly by duplicate meals. The oxalate and creatinine content of ten successive 24-hour urine samples were then measured. It was soon apparent that the daily urinary oxalate excretion fluctuates appreciably in individuals even under closely controlled conditions of diet and exercise and that even subjects initially comparable may differ in their fluctuation responses. The average excretion of oxalate in these circumstances was equivalent to 22 (S.D.7) mg. $(COOH)_2 . 2H_2O$ per 24 hours. [illegible] the average [illegible] creatinine, [illegible] This total [illegible] Lamden and [illegible] (S.D. 20·2) [illegible] ere working under more closely controlled conditions and the methods of analysis not identical.

From further analyses of many urinary samples over several years, we have not discovered any deviation from our normal range and if due allowance is made for the small experimental loss the total oxalate excretion in normality does not exceed 40 mgm. per 24 hours.

A further series of volunteers under similar conditions of control and diet, were then given sodium oxalate in solution by mouth after each meal. In each case they responded with an increase in urinary oxalate. The increase varied between

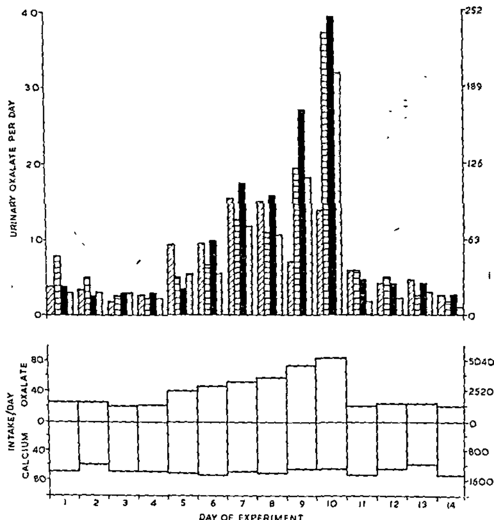


FIG 16 The effect of oral ingestion of sodium oxalate on the urinary oxalate excretion by four normal subjects

2·3 per cent and 4·5 per cent of the ingested dose, and there was an immediate return to the basic level when the oxalate administration ceased (Fig. 16). Analysis of the individual figures leads us to expect that on average the dietary intake of soluble

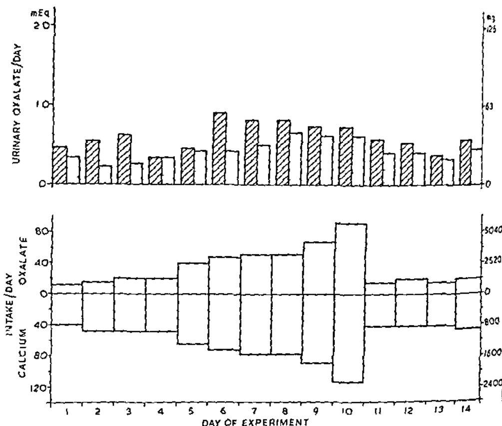


FIG. 17. The effect of oral ingestion of calcium oxalate on the urinary oxalate excretion by four normal subjects.

oxalate would have to increase at least two and a half-fold to produce any significant rise in urinary oxalate. To attain the oxalate levels of primary hyperoxaluria would be virtually impossible by dietetic measures and could only be done by soluble oxalate administration and even then the increase is transient.

This experiment was then repeated using equimolar quantities of calcium instead of sodium oxalate. To avoid possible error this test was only carried out in two volunteers, in whom we had ascertained a normal gastric acidity. In neither case was there any significant increase in the urinary oxalate content (Fig. 17).

The extensive literature on oxalate intake as an aetiological factor in urinary calculus formation was reviewed by Jeghers and Murphy (1945). Much of this work remains, at best, pure hypothesis and some is invalidated by the absence of previous information on the urinary oxalate content.

It is apparent that the water-soluble oxalate content of the diet might be expected to influence the urinary oxalate excretion, but when the enormous doses in our experiments are considered in relation to the relatively small changes that can occur from food, it is unlikely that alterations in dietary oxalate can ever be held responsible for material changes in the oxalate excretion in the urine. In addition, even if a sudden disproportionate amount of the few foods rich in soluble oxalate is ingested the increased oxalate output is very short-lived.

Metabolic Studies in Primary Hyperoxaluria

A study of the available evidence, clinical, pathological, genetic and chemical, leads to the supposition that the essential abnormality in this condition was probably a metabolic error (Archer, Dormer, Scowen and Watts, 1958a). It was, however, first necessary to ascertain whether subjects with primary hyperoxaluria behaved normally to an increase in oxalate load and to try and establish that an error producing either excessive intestinal absorption or an increased renal leak of oxalate was not present, before proceeding to investigate the more precise metabolic disturbance.

Two patients with primary hyperoxaluria were given a standard diet as had previously been used in the normal controls. Sodium and later calcium oxalate were then given in increasing dosage, the total dose being far in excess of that administered to normals. The sodium oxalate was given in solution and the calcium oxalate in cachets, both immediately

2·3 per cent and 4·5 per cent of the ingested dose, and there was an immediate return to the basic level when the oxalate administration ceased (Fig. 16). Analysis of the individual figures leads us to expect that on average the dietary intake of soluble

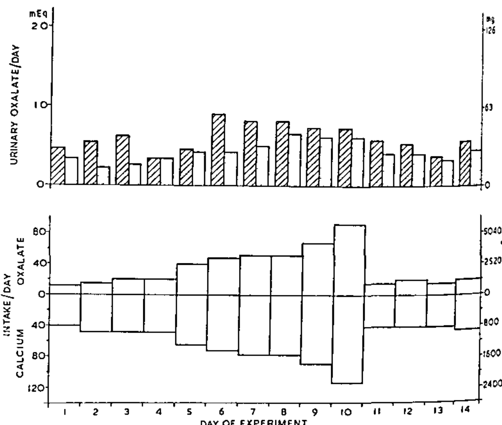


Fig. 17. The effect of oral ingestion of calcium oxalate on the urinary oxalate excretion by four normal subjects

oxalate would have to increase at least two and a half-fold to produce any significant rise in urinary oxalate. To attain the oxalate levels of primary hyperoxaluria would be virtually impossible by dietetic measures and could only be done by soluble oxalate administration and even then the increase is transient.

This experiment was then repeated using equimolar quantities of calcium instead of sodium oxalate. To avoid possible error this test was only carried out in two volunteers, in whom we had ascertained a normal gastric acidity. In neither case was there any significant increase in the urinary oxalate content (Fig. 17).

The extensive literature on oxalate intake as an aetiological factor in urinary calculus formation was reviewed by Jeghers and Murphy (1945). Much of this work remains, at best, pure hypothesis and some is invalidated by the absence of previous information on the urinary oxalate content.

It is apparent that the water-soluble oxalate content of the diet might be expected to influence the urinary oxalate excretion, but when the enormous doses in our experiments are considered in relation to the relatively small changes that can occur from food, it is unlikely that alterations in dietary oxalate can ever be held responsible for material changes in the oxalate excretion in the urine. In addition, even if a sudden disproportionate amount of the few foods rich in soluble oxalate is ingested the increased oxalate output is very short-lived.

Metabolic Studies in Primary Hyperoxaluria

A study of the available evidence, clinical, pathological, genetic and chemical, leads to the supposition that the essential abnormality in this condition was probably a metabolic error (Archer, Dormer, Scowen and Watts, 1958a). It was, however, first necessary to ascertain whether subjects with primary hyperoxaluria behaved normally to an increase in oxalate load and to try and establish that an error producing either excessive intestinal absorption or an increased renal leak of oxalate was not present, before proceeding to investigate the more precise metabolic disturbance.

Two patients with primary hyperoxaluria were given a standard diet as had previously been used in the normal controls. Sodium and later calcium oxalate were then given in increasing dosage, the total dose being far in excess of that administered to normals. The sodium oxalate was given in solution and the calcium oxalate in cachets, both immediately

after food. Finally freshly expressed rhubarb juice was given to ascertain whether the oxalate excretion was proportional to the oxalate content alone, which was estimated in an aliquot of the expressed juice.

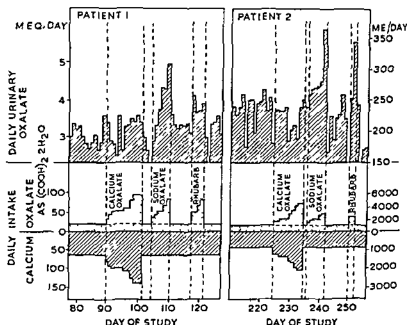


FIG. 18 The changes in the daily urinary oxalate excretion with the administration of sodium and calcium oxalate and expressed rhubarb juice of two patients with primary hyperoxaluria.

The increase in excretion with sodium oxalate closely resembled the normal pattern both in time and percentage of administered dose. Calcium oxalate, as in the normal subjects, did not materially alter the oxalate output. The excretion of additional oxalate which occurred after the ingestion of rhubarb juice did not exceed that which would be expected on the basis of its oxalate content (Fig. 18). It seemed unlikely in view of these findings that either excessive absorption or renal leak could be held responsible for the hyperoxaluria.

A study of the possible mechanisms by which excessive oxalate formation might be produced leads us to postulate that an

unusual conversion of glycine to oxalate might be one of the errors responsible (Fig. 19).

An endeavour was therefore made in two patients with primary hyperoxaluria to reduce the size of the glycine pool by

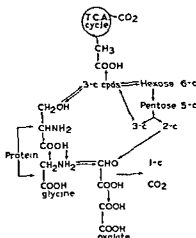


FIG. 19. Possible metabolic pathways in the conversion of glycine to oxalate.

protein starvation. It was difficult to ascertain with certainty that this could happen and for practical reasons not possible to continue the protein starvation for inordinate lengths of time. Nevertheless by gross protein reduction in the diet over three-week periods, there occurred a small but statistically significant fall in the oxaluria. This was contrasted with similar periods with a fixed protein intake of average normal figures. Repeated observations of these periods showed no statistical significance in the fluctuation. This evidence suggested, but did not seem conclusive of, the fact that protein depletion by diminishing the amino-acid pools, perhaps glycine, might be a factor in influencing the oxaluria. Nevertheless it appeared encouraging enough for us to persevere with our earlier hypothesis.

Two further lines of inquiry were then pursued (Archer, Dormer, Scowen and Watts, 1958a). First an experiment was undertaken to increase the glycine load by adding increasing

quantities of glycine to the diet. A female patient with hyperoxaluria, aged 12, was used for this experiment. We had little knowledge of how much we could give and what effects might be expected and therefore began by observing the effects in a 12-year-old normal control.

The daily oxalate output in the control over four days before and after the glycine administration was 4 to 10 mgm., and 8 to 12 mgm. of oxalate as oxalic acid dihydrate. Glycine was then given from 20 G. to 100 G. daily over six days in four equal and equally spaced doses. There was some decline in appetite, but no other ill effects. The urinary oxalate as oxalic acid dihydrate fluctuated between 2 and 12 mgm. per 24 hours. It was clear that a great increase of glycine could be tolerated in the normal child and that no increase in oxalate output would occur.

When an endeavour was made to repeat this experiment in the patient with hyperoxaluria it did not prove as easy. She was given initially 4·5 G. glycine 6-hourly with the object of increasing the dose to 50 or 100 G. if possible. As soon as the dose exceeded 20 G. per 24 hours, nausea and occasional vomiting occurred and little additional diet or fluid were taken. The dose could not be raised above 50 G. per 24 hours and after 8 days the administration ceased, because of extensive lethargy, nausea and vomiting. There occurred an appreciable rise in the oxaluria, but there also occurred a marked increase in amino-aciduria, analysis of which showed a gross glycinuria. In spite of vomiting which made the ingested dose uncertain, there had clearly occurred a great increase in glycine absorption overall; this had produced a glycinuria and a rise in the urine oxalate (Fig. 20). These findings contrasted sharply with the normal, but did not necessarily implicate the glycine, as the associated illness with vomiting might have precipitated the increase in oxalate by some other means. It was not, however, easy to proceed with this type of experiment for the obvious practical reasons, nor did we feel that it would produce more direct information.

We decided to try and deplete the glycine pool by the administration of benzoate, thus removing large amounts of

PLATE XX





FIG. 1 X-rays of renal tract to show the rapid recurrence of urinary calculi over the course of three years in a fatal case of primary hyperoxaluria.

quantities of glycine to the diet. A female patient with hyperoxaluria, aged 12, was used for this experiment. We had little knowledge of how much we could give and what effects might be expected and therefore began by observing the effects in a 12-year-old normal control.

The daily oxalate output in the control over four days before and after the glycine administration was 4 to 10 mgm., and 8 to 12 mgm. of oxalate as oxalic acid dihydrate. Glycine was then given from 20 G. to 100 G. daily over six days in four equal and equally spaced doses. There was some decline in appetite, but no other ill effects. The urinary oxalate as oxalic acid dihydrate fluctuated between 2 and 12 mgm. per 24 hours. It was clear that a great increase of glycine could be tolerated in the normal child and that no increase in oxalate output would occur.

When an endeavour was made to repeat this experiment in the patient with hyperoxaluria it did not prove as easy. She was given initially 4·5 G. glycine 6-hourly with the object of increasing the dose to 50 or 100 G. if possible. As soon as the dose exceeded 20 G. per 24 hours, nausea and occasional vomiting occurred and little additional diet or fluid were taken. The dose could not be raised above 50 G. per 24 hours and after 8 days the administration ceased, because of extensive lethargy, nausea and vomiting. There occurred an appreciable rise in the oxaluria, but there also occurred a marked increase in amino-aciduria, analysis of which showed a gross glycinuria. In spite of vomiting which made the ingested dose uncertain, there had clearly occurred a great increase in glycine absorption overall; this had produced a glycinuria and a rise in the urine oxalate (Fig. 20). These findings contrasted sharply with the normal, but did not necessarily implicate the glycine, as the associated illness with vomiting might have precipitated the increase in oxalate by some other means. It was not, however, easy to proceed with this type of experiment for the obvious practical reasons, nor did we feel that it would produce more direct information.

We decided to try and deplete the glycine pool by the administration of benzoate, thus removing large amounts of

PLATE XXII





FIG. 3 Kidney showing calcium oxalate crystals in *scarred* glomerulus in tubules and in interstitial tissues. Half-crossed Nicol prisms H & E (×125).
On the right von Kossa (×125)



FIG 2 Kidney showing calcium oxalate crystals most of which are related to the remains of renal tubules Much interstitial fibrosis Half-crossed Nicol prisms. Haematoxylin and eosin (×50)

PLATE XXIII





FIG 4 (*left*) Myocardium calcium oxalate crystals in the connective-tissue septa some lie directly over, or within, muscle fibres Half-crossed Nicol prisms H & E ($\times$115)

FIG 5 (*right*) Calcium oxalate crystals in the myocardium Crossed Nicol prisms ($\times$650).

PLATE XXVI

FIG. 10 Testis, calcium oxalate crystals within lumen and in epithelial lining of rete testis. Seminiferous tubules also contain crystals Half-crossed Nicol prisms H & E (×50)



FIG 9 Splenic arteriole with intramural calcium oxalate crystals Half-crossed Nicol prisms H & E (×120)



FIG. 11 Femoral bone marrow; single calcium oxalate crystal closely related to a trabecula Half-crossed Nicol prisms. H & E (×183).

Plate XXV

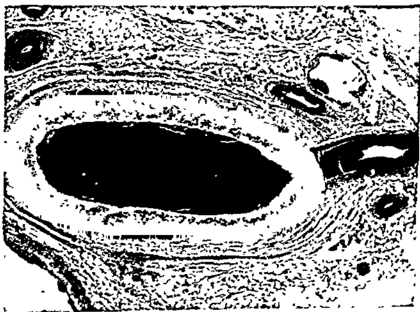


Fig 7 Femoral artery showing extensive oxalate deposit in the muscular tunica media Half-crossed Nicol prisms.

Fig 8. Carotid artery showing no crystalline deposit. Half-crossed Nicol prisms.

glycine by the combination to produce hippuric acid. Two normal subjects, one adult and one child, were given sodium benzoate by mouth in increasing doses. The child had an initial daily dose of 5 G. rising to 12 G. in divided doses. The adult

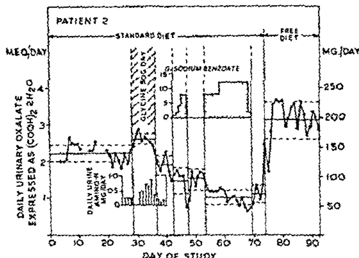


Fig. 20. The daily urinary oxalate excretion on a patient with primary hyperoxaluria during periods of glycine and sodium benzoate administration, together with the results obtained during the corresponding control periods. The average value for each experimental and control period is represented by an uninterrupted horizontal line, ± 1 standard deviation about each mean value is indicated by the interrupted horizontal lines.

dose began at 10 G. and rose to 40 G. per 24 hours. The urinary oxalate was measured over a prolonged control period before the experiment began and continued during and after the benzoate administration. Though large amounts of hippuric acid were excreted no significant change occurred in oxalate output.

It was of interest to note that the adult in the dose range between 30 and 40 G. of benzoate developed a marked malaise, gross impairment of concentration and sensations of remoteness amounting to considerable disorientation.

Two patients with hyperoxaluria, one adult and one child, of the same sex, approximately the same age as the controls,

PLATE XXVII

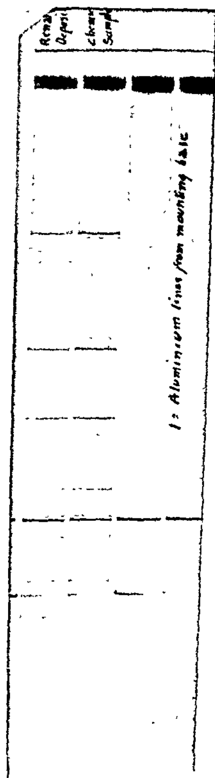


FIG. 12 X-ray diffraction spectra of the calcium oxalate present in the kidney of a case of primary hyper-oxaluria and an authentic example of calcium oxalate monohydrate (Analysis performed by Dr R J R Cureton).

glycine by the combination to produce hippuric acid. Two normal subjects, one adult and one child, were given sodium benzoate by mouth in increasing doses. The child had an initial daily dose of 5 G. rising to 12 G. in divided doses. The adult

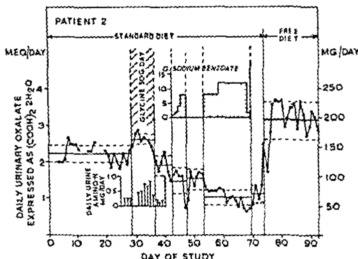


FIG 20 The daily urinary oxalate excretion on a patient with primary hyperoxaluria during periods of glycine and sodium benzoate administration, together with the results obtained during the corresponding control periods. The average value for each experimental and control period is represented by an uninterrupted horizontal line, $\pm$ 1 standard deviation about each mean value is indicated by the interrupted horizontal lines

dose began at 10 G. and rose to 40 G per 24 hours. The urinary oxalate was measured over a prolonged control period before the experiment began and continued during and after the benzoate administration. Though large amounts of hippuric acid were excreted no significant change occurred in oxalate output.

It was of interest to note that the adult in the dose range between 30 and 40 G. of benzoate developed a marked malaise, gross impairment of concentration and sensations of remoteness amounting to considerable disorientation.

Two patients with hyperoxaluria, one adult and one child, of the same sex, approximately the same age as the controls,

were treated similarly with sodium benzoate. This was associated with a marked decrease in oxalate output in both. In the adult it was less well marked than in the child, but it was impossible to maintain the higher doses in the adult owing to the development of marked psychological disturbance identical with that described in the normal adult control. It will be seen from Figure 20 that in the child the excretion of oxalate was decreased to a high normal range, but cessation of benzoate administration was followed by a rapid relapse of the hyperoxaluria. This followed a short period of free diet owing to the development of renal colic and it is possible that subsequent high figures may be aggravated by the passage of oxalate gravel, though care was taken to exclude small calculi which were passed at this time.

An attempt was made to lower the oxalate output more permanently by prolonged administration of benzoate, but was unsuccessful. The excretion fell again, but after four weeks began to rise and in spite of the benzoate gradually returned to the hyperoxaluric levels. It is of interest that each time the urinary oxalate fell to near normal, renal colic and passage of small oxalate calculi occurred.

Although we had no absolute proof it seemed that the changes in oxalate excretion with the administration of sodium benzoate provided some experimental proof that glycine might be one, if not the main precursor of the excessive oxalate output and it was probable that if this were so it was occurring because there was an excessive glyoxylate load from the abnormal glycine pathway (see Fig. 19).

It was possible that the apparent absence of effect of benzoate in influencing the normal oxalate output might be misleading. For a small variation would not be observable in the light of apparently spontaneous fluctuations in the normal (Archer, Dormer, Scowen and Watts, 1957b). A partial origin from glycine could not be excluded therefore by these experiments. Sodium glyoxalate was therefore administered to a normal adult with a mean normal oxalate excretion of 23 mgm. per 24 hours. This subject was given 48 0 m. mol. of sodium glyoxalate by mouth divided in doses over three days (0 228 m. mol. per

kg.). This resulted in a gross hyperoxaluria, the rise exceeding 200 mgm. per 24 hours. Clinically this was associated with considerable malaise and lethargy with nausea but no other specific symptoms. The proportion of ingested glyoxalate excreted as oxalate was 8·42 per cent.

An adolescent female patient with hyperoxaluria with a mean daily excretion of 159 mgm. of oxalate was then given 9·07 m. mol. of sodium glyoxalate in increasing doses to reach 0·227 m. mol. per kg. Again there was an immediate rise in the oxalate output to over 380 mgm. per 24 hours and the total excretion of ingested glyoxalate as oxalate was 23·9 per cent.

Clearly an increased glyoxalate load will induce hyperoxaluria in the normal and aggravate the primary hyperoxaluria. The greater proportion in the primary hyperoxaluria would be in line with the previously postulated error.

Glycollic aldehyde and sodium glycollate were then administered to the same patient in approximately equimolecular quantities. The oxalate output during the administration of these substances did not fluctuate beyond the standard deviation of the prolonged central observations (Fig. 21).

It seemed probable from these observations that the essential metabolic lesion probably resulted from an excessive glyoxalate formation. Since it appeared that the glycollic aldehyde pathway (see Fig. 19) did not materially increase the oxalate output the important link was the reaction sequence of glycine to glyoxalate. This was in conformity with the earlier postulate, and in the absence of methods to establish with certainty the oxalate and glyoxalate content of the blood, it became necessary to demonstrate, if possible, the direct conversion of glycine to oxalate in man and the excessive conversion in hyperoxaluria. We were not willing to expose our young patients to ^{14}C and therefore made use of the stable isotope ^{13}C. For this purpose ^{13}C-labelled glycine was prepared (Watts and Crawhall, 1959).

In order to use this isotope it was necessary to ascertain the extent of the dilution which a given dose would undergo. To express the results an estimate had to be made of the size and turnover of the glycine pool. The first glycine metabolic pool was used (Arnstein and Neuberger, 1951). It represents the sum

*were treated similarly with sodium benzoate. This was associ*ated with a marked decrease in oxalate output in both. In the adult it was less well marked than in the child, but it was *impossible to maintain the higher doses in the adult owing to* the development of marked psychological disturbance identical with that described in the normal adult control. It will be seen *from Figure 20 that in the child the excretion of oxalate was* decreased to a high normal range, but cessation of benzoate administration was followed by a rapid relapse of the hyper*oxaluria. This followed a short period of free diet owing to the* development of renal colic and it is possible that subsequent high figures may be aggravated by the passage of oxalate gravel, though care was taken to exclude small calculi which were passed at this time.

An attempt was made to lower the oxalate output more permanently by prolonged administration of benzoate, but was unsuccessful. The excretion fell again, but after four weeks began to rise and in spite of the benzoate gradually returned to the hyperoxaluric levels. It is of interest that each time the urinary oxalate fell to near normal, renal colic and passage of small oxalate calculi occurred.

Although we had no absolute proof it seemed that the changes in oxalate excretion with the administration of sodium benzoate provided some experimental proof that glycine might be one, if not the main precursor of the excessive oxalate output and it was probable that if this were so it was occurring because there was an excessive glyoxylate load from the abnormal glycine pathway (see Fig. 19).

It was possible that the apparent absence of effect of benzoate in influencing the normal oxalate output might be misleading. For a small variation would not be observable in the light of apparently spontaneous fluctuations in the normal (Archer, Dormer, Scowen and Watts, 1957b). A partial origin from glycine could not be excluded therefore by these experiments. Sodium glyoxalate was therefore administered to a normal adult [illegible] 24 hours. [illegible] glyoxalate by mouth [illegible] *mol. per*

total of uncombined visceral glycine, the glycine in the plasma and interstitial fluid. A dose of glycine by mouth would be partially utilized in the liver, for hippuric acid synthesis, the remainder would pass into the blood and mingle in the first

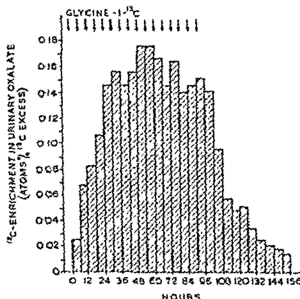


FIG 22 The ^{13}C content of the calcium oxalate isolated from successive 6-hour collections in a case of primary hyperoxaluria given glycine-1-^{13}C at the times indicated by the arrows.

glycine pool. On the basis of equilibrium in this pool the urinary glycine components should represent the plasma equilibrium as between isotopically labelled and other glycine. Studies by Watts and Crawhall (1959) have shown the validity of this procedure and made possible the technique of multiple-dose administration. By this means it is possible to measure the isotope content in 6-hour urine collections.

The first report of this method was made in 1958 (Scowen, Crawhall and Watts, 1958). Glycine-1-^{13}C was given orally to a patient with primary hyperoxaluria (2·56 mgm. per kg., 49·8 atoms per cent ^{13}C excess 6-hourly for 4 days). The oxalate

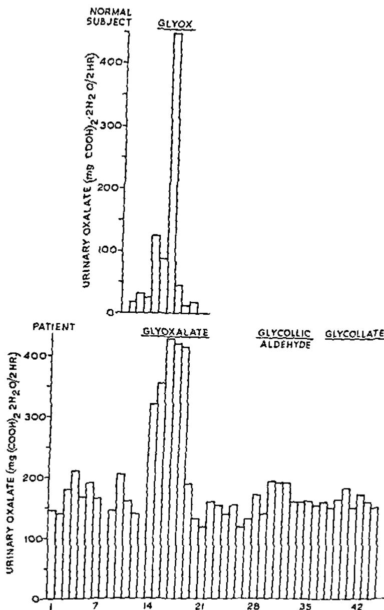


FIG. 21. The effect of the administration of sodium glyoxalate on the urinary oxalate excretion of a normal subject, and of sodium glyoxalate, glycollic aldehyde and sodium glycollate on the urinary oxalate excretion of a patient with primary hyperoxaluria

total of uncombined visceral glycine, the glycine in the plasma and interstitial fluid. A dose of glycine by mouth would be partially utilized in the liver, for hippuric acid synthesis, the remainder would pass into the blood and mingle in the first

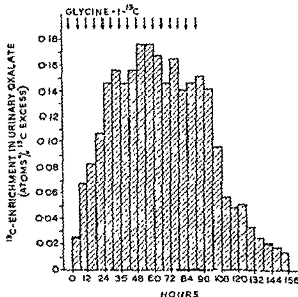
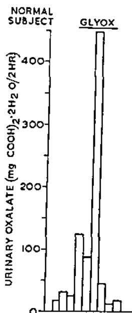
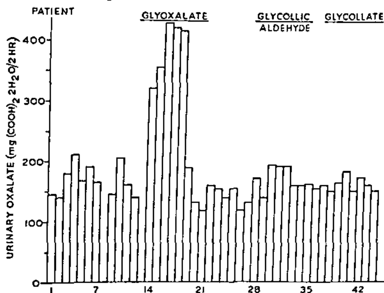


Fig 22 The ^{13}C content of the calcium oxalate isolated from successive 6-hour collections in a case of primary hyperoxaluria given glycine -1-^{13}C at the times indicated by the arrows

glycine pool. On the basis of equilibrium in this pool the urinary glycine components should represent the plasma equilibrium as between isotopically labelled and other glycine. Studies by Watts and Crawhall (1959) have shown the validity of this procedure and made possible the technique of multiple-dose administration. By this means it is possible to measure the isotope content in 6-hour urine collections.

The first report of this method was made in 1958 (Scowen, Crawhall and Watts, 1958). Glycine-1-^{13}C was given orally to a patient with primary hyperoxaluria (2·56 mgm. per kg., 49·8 atoms per cent ^{13}C excess 6-hourly for 4 days). The oxalate


NORMAL SUBJECT
GLYOX
URINARY OXALATE (mg COOH)$_2$·2H$_2$O/2HR)
400
300
200
100
0


PATIENT
GLYOXALATE
GLYCOLLIC ALDEHYDE
GLYCOLLATE
URINARY OXALATE (mg (COOH)$_2$ 2H$_2$O/2HR)
400
300
200
100
0
1
7
14
21
28
35
42

total of uncombined visceral glycine, the glycine in the plasma and interstitial fluid. A dose of glycine by mouth would be partially utilized in the liver, for hippuric acid synthesis, the remainder would pass into the blood and mingle in the first

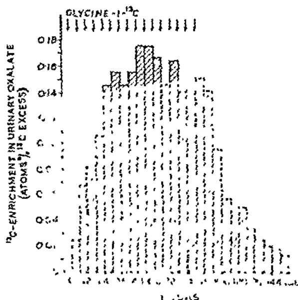


FIG 22 The ^{13}C content of the calcium oxalate isolated from successive 6-hour collections in a case of primary hyperoxaluria given glycine-1-^{13}C at the times indicated by the arrows

glycine pool. On the basis of equilibrium in this pool the urinary glycine components should represent the plasma equilibrium as between isotopically labelled and other glycine. Studies by Watts and Crawhall (1959) have shown the validity of this procedure and made possible the technique of multiple-dose administration. By this means it is possible to measure the isotope content in 6-hour urine collections.

The first report of this method was made in 1958 (Scowen, Crawhall and Watts, 1958) Glycine-1-^{13}C was given orally to a patient with primary hyperoxaluria (2 56 mgm. per kg., 49 8 atoms per cent ^{13}C excess 6-hourly for 4 days). The oxalate

was isolated from successive 6-hour collections and its ^{13}C content measured. There was prompt and marked incorporation of the isotope in the urinary oxalate (Fig. 22).

Comparison was then made with the isotope enrichment of

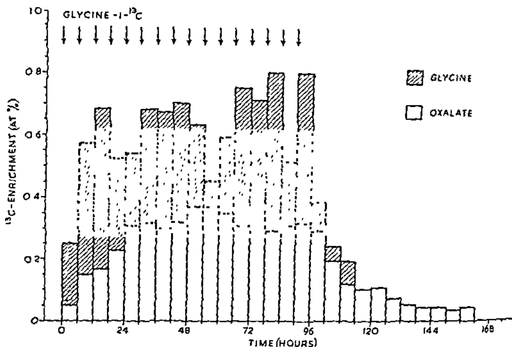


FIG. 23. The ^{13}C content of the urinary oxalate and of the urinary free glycine excreted by a patient with primary hyperoxaluria given glycine -1-^{13}C at the times indicated by the arrows

the urinary free glycine, which as has already been stated represents a valid sample of the free glycine metabolic pool. It is interesting to observe (Fig. 23) the close relationship between the comparable figures.

We have since repeated these observations on two further patients with primary hyperoxaluria with nearly identical results.

By calculation it seems probable that a major part of the hyperoxaluria can be attributed to this pathway from glycine. It does not account for the whole: this still requires further

elucidation, and it may well be that other substances are involved. There have been grounds for suggesting that ascorbic acid may be converted to oxalate in man as well as animals. This was shown after administration of ascorbic acid by Lamden and Chrystowski (1953, 1954) and more precisely by Hellman and Burns (1958) using L-ascorbic acid-l-^{14}C, though the quantity of oxalate produced by this route seems limited.

Although much work remains to be done on these other pathways, it now appears established that the major pathway of oxalate is from glycine via glyoxalate; it remains to be discovered whether the failure lies in the degradation of glyoxalate or whether there is excessive, possibly abnormal conversion of glycine to glyoxalate.

From the preliminary observations it would seem that in the normal it is easy to overload the glyoxalate combustion mechanism. Accumulation of glyoxalate will readily induce hyperoxaluria. The available evidence is in conformity with the hypothesis that excessive glyoxalate is formed from glycine amongst other substances with resultant hyperoxaluria, rather than a breakdown in glyoxalate combustion to carbon dioxide and water.

In this connection we have recently been able to show that in a normal subject given ^{13}C-labelled glycine, some of the urinary oxalate is produced from this glycine, which suggests that this is a normal pathway of metabolism and the conversion greatly increased in primary hyperoxaluria.

Summary

The term hyperoxaluria is defined. Hyperoxaluria may occur in a number of conditions including the ingestion of excessive soluble oxalate or its precursors. Such conditions produce transient hyperoxaluria. A condition of persistent hyperoxaluria is described unrelated to exogenous circumstances, congenital in origin and termed primary hyperoxaluria. The clinical and morbid anatomical features of this entity are described and the diffuse deposition of oxalate in the body defined as oxalosis. The evidence suggests that this morbid anatomical feature is probably a terminal manifestation. Familial studies show that the

abnormality is inherited and the mode of inheritance is consistent with the operation of a rare recessive character.

Investigation of the condition itself indicates a metabolic error. There is an abnormal conversion from glycine to oxalate, probably through an excessive glyoxalate formation. This metabolic error accounts for a great part of the increased oxalate excretion; other errors, however, await further investigation to account for the whole.

The congenital nature of the disease, its mode of inheritance and the demonstration of the metabolic error lead us to designate this entity as an inborn error of metabolism.

REFERENCES

APONTE, G. E and FETTER, T. R (1954). *Am. J. Clin. Path.* 24, 1363

ARCHER, H E, DORMER, A E., SCOWEN, E. F. and WATTS, R. W. E. (1957a). *Lancet*, ii, 320

ARCHER, H. E., DORMER, A. E., SCOWEN, E. F. and WATTS, R. W. E. (1957b). *Clin. Sci* 16, 405

ARCHER, H. E., DORMER, A. E., SCOWEN, E. F. and WATTS, R. W. E. (1958a). *Brit med. J.* i, 175.

ARCHER, H. E., DORMER, A. E., SCOWEN, E. F. and WATTS, R W. E. [illegible]

[illegible] *J* 50, 154.

BURKE, E. C., BAGGENSTOSS, A H, OWEN, CH. A., JR., POWER, M. H. and LOHR, O W. (1955) *Paediatrics*, 15, 383

CARSON, M. J. (1951) *J. Pediat.* 39, 251.

CHOU, L. Y and DONOHUE, W. I. (1952). *Paediatrics*, 10, 660.

CRAWHALL, J C and WATTS, R W. E (1958). *Biochem. J.* 69, 2 22 P.

DAVIS, J. S., KLINGBERG, W G. and STOWELL, R E. (1950). *J Pediat* 36, 323.

DUNN, H G. (1955). *Am. J. Dis Child* 90, 58

EDWARDS, D L (1957). *Archiv. Path* 64, 546.

GRAM, H. C (1932) *Acta med. scand* 78, 268.

HELLMAN, D. and BURNS, J. J. (1958) *J biol Chem.* 230, 923

[illegible]

233, 208

34, 173, 60.

[illegible] 12, 420.

LAMDEN, M. P. and CHRYSTOWSKI, G. A. (1954). *Proc. Soc Exp Biol N.Y.* **85**, 190.

LEPOUTRE (1925). *J. Urol. méd. chir* **20**, 424.

[illegible] **38**, 353.

[illegible] *ch. Franç. Pediatric.* **14**, 360.

NEUSTEIN, H. B , STEVENSON, S. S and KRAINER, L. (1955). *J. Pediat.* **47**, 624

NEWNS, G. H and BLACK, J. A. (1953). *Great Ormond St. J.* **40**

ÖIGAARD, H. and SODERHJELM, L. (1957). *Acta soc. med. upsaliensis* **62**, 176.

OSTRY, H (1951). *Canad M.A.J.* **65**, 465

POWERS, H. H , LEVATIN, P. (1944). *J. biol Chem.* **154**, 207

SCOWEN, E. F., CRAWHALL, J C. and WATTS, R. W. E. (1958). *Lancet*, **ii**, 300

SCOWEN, E. F , STANSFELD, A G. and WATTS, R. W E (1959). *J. Path. Bact.* **77**, 195

SCOWEN, E. F , WATTS, R. W. E. and HALL, E. G. (1959). *Annals Human Genetics* In press.

SHEPHARD, T. H., KREBS, E. C. and LEE, L S (1958). *Amer. J. Dis. Child.* **96**, 490

SIMKÓ, I (1957) *Ann Paediat* **189**, 1.

VAUGHAN, J H , SOSMAN, M. C. and KINNEY, T. D (1947). *Am. J. Roentgenol* **58**, 33.

VISCHER, W. (1947) *Schweiz. Z. Allgem. Path* **10**, 286.

WATTS, R W E and CRAWHALL, J. C. (1959). In press.

WOHL, H. (1942) *J. Pediat.* **21**, 382

ZOLLINGER, H. U and ROSENMUND, H (1952) *Schweiz. Med Wochschr.* **82**, 1261.

XXI

Sodium Excretion and the Control of Extracellular Fluid Volume

OLIVER WRONG

ALMOST every schoolboy knows that man is about 60 per cent water. Perhaps we learn this fact too early in our education, for we soon push it to the back of our minds as old stuff, and subsequently fail to recognize many of its important implications. We must remind ourselves, for example, that body water is somehow controlled with great accuracy, for in health, despite a large and variable fluid intake, the weight of the body varies by extremely little, plus or minus 1 per cent, from day to day. This lecture is intended to discuss how this constancy is achieved, with particular emphasis on that portion of body water which lies outside the cells.

Intracellular Fluid

Brief mention must first be made of the two-thirds of body water which lies within the cells and yet exerts important influences on extracellular fluid volume. Little is known about mammalian cell fluid, for it is difficult to obtain and alters rapidly in composition when cellular metabolism is interrupted. Consequently there has been disagreement about what, to many people, must seem one of the most fundamental characteristics of this fluid—its osmotic strength. Many workers of the thirties and forties assumed that intracellular and extracellular fluids were in osmotic equilibrium, and this assumption was supported by the behaviour of the erythrocyte (admittedly rather an atypical cell) *in vitro*, and the internal shifts of body water which

could be demonstrated after osmotic distortions of extracellular fluid (Darrow and Yannet, 1935). As the concentration of known cation in cells is about 40 per cent higher than in extracellular fluid it was necessary to postulate, in order to satisfy the demands both of electroneutrality and of osmotic equilibrium, either that the majority of intracellular anions were polyvalent (a reasonable supposition in view of the amounts of phosphate and protein contained in cells) or that a portion of intracellular cation was bound in osmotically inactive form.

But the assumption of intracellular isotonicity, although it fitted in with most of the observed facts, remained until very recently only an assumption and could not be confirmed by direct experiment. Indeed, as long ago as 1901 Sabbatani attempted to determine intracellular osmolality by measuring the freezing-points of tissues, and recorded osmolalities of one and a half or more times that of extracellular fluid, a finding which has since been confirmed by other workers. On the other hand Conway and McCormack (1953), also using freezing-point determinations, found intracellular osmolalities similar to that of extracellular fluid, and attributed the high figures of earlier workers to rapid autolysis of cell constituents prior to freezing.

cells swell *in vitro* when suspended in isotonic saline media, and Robinson demonstrated that this swelling was more pronounced when cellular metabolism was inhibited by cold, cyanide, or oxygen lack Opie and Robinson attributed this cellular swelling to breakdown in an active process of water extrusion, and they postulated that the cell is normally hypertonic to extracellular fluid, maintaining its volume by this continual extrusion of water. However, both Mudge (1951) and Leaf (1956) have shown that the fluid which enters cells when their metabolism is disturbed is not water alone, but a solution of sodium and chloride approximately isotonic with the ambient medium Leaf has pointed out that this finding does not imply cellular hypertonicity, for the swelling can be explained as a natural consequence of breakdown in active cellular extrusion of sodium

—the osmotic effect of non-diff...
then lead to entry of isot...
the advantages of this scheme is that active transport of sodium has been demonstrated in other situations (Hodgkin and Keynes, 1955; Leaf, Anderson and Page, 1958) whereas there is no certain model for the active transport of water. Within the last few months Maffly and Leaf (1958) have reported new determinations of intracellular osmolality, using an ingenious technique for measuring the *melting-point* of instantaneously frozen tissues, and so avoiding all possibility of autolysis. Muscle, liver, and brain all gave values identical with that of extracellular fluid.

The problem of intracellular tonicity thus appears to be resolved, and we may reasonably conclude that the mammalian cell is in equilibrium with its extracellular environment (although this generalization may not apply to cells which elaborate hypotonic or hypertonic secretions, such as those of the salivary and sweat glands and renal tubules). The careful clinical studies of Wynn (1957), and of Edelman, Leibman, Omeara and Birkenfeld (1958), suggest that this generalization can be extended to cover the disturbances of electrolyte and water metabolism which occur in disease. This conclusion has far-reaching implications which justify this digression into the intracellular space. For, if the intracellular compartment as a whole behaves as a perfect osmometer, in osmotic equilibrium with extracellular fluid, the following deductions are valid:

(1) *Change in extracellular osmolality will cause shifts of water across the cell membrane and resultant change in intracellular osmolality and volume.* For example, hypernatraemia, from water deficit, will cause shrinking and concentration of the intracellular...
the hy...

(2) On the other hand *change in extracellular volume alone will not affect either intracellular volume or osmolality.* Perhaps it is for this reason that the body responds more slowly to changes in extracellular volume than to changes in solute concentration.

(3) *Change in extracellular volume may result from change in intracellular osmolality* (the reverse of (1) above). For example, the hyponatraemia of some patients with heart failure may be partly the result of an intracellular potassium deficiency, and can be corrected by administration of potassium salts (Cort and Matthews, 1954; Laragh, 1954).

(4) Provided intracellular osmolality remains unchanged, *intracellular volume is proportional to total intracellular solute, chiefly salts of potassium.* When potassium is deficient it can be partially replaced by sodium and hydrogen ion (Cooke, Segar, Cheek, Coville and Darrow, 1952) and the basic amino-acid lysine (Eckel, Norris and Pope, 1958). But in normal circumstances intracellular volume must depend largely on intracellular potassium content, and there is practically no information on how the latter is controlled; nor at present is there much knowledge of how the kidney regulates the total amount of potassium in the body.

Extracellular Fluid

The extracellular fluid extends to every crevice of the body, much of it occupying lax and easily distensible connective tissue. Except for the intravascular component, which is enclosed by the relatively structureless vascular endothelium, it is not confined by any specialized system of cells, and consequently it is very difficult for us to envisage any mechanism whereby the body could become aware of changes in the volume of this fluid. Yet certainly the volume is controlled within narrow limits, which implies some sort of 'awareness', and it is to our advantage that this is so. An inadequate volume leads to reduction in the vital intravascular portion, with hypotension and that combination of clinical signs that we call 'dehydration', whereas an excessive volume is liable to collect as oedema, interfering with the transport of oxygen and metabolites between capillary and cell, and threatening asphyxia if it accumulates in the lungs.

The control of extracellular volume has been rather neglected by physiologists, who usually prefer to think in terms of concentrations within a fluid, rather than the volume of the fluid itself. Yet now and again some of the most distinguished workers

in the field have given their attention to the subject. Claude Bernard, who originated our concept of extracellular fluid as an entity, specifically mentioned control of volume: 'Chez les animaux à vie libre il doit exister un ensemble de dispositions réglant les pertes et les apports de manière à maintenir la quantité d'eau nécessaire dans le milieu intérieur.' (Bernard, 1878.)

Thirty years later Starling (1909) recognized that intestinal absorption of salt and water proceeds regardless of the needs of the body, and pointed out that control of extracellular fluid volume must therefore devolve on the kidneys. But in spite of these auspicious beginnings the early part of the century saw an almost complete lack of speculation and experiment on this important subject. Instead, in the twenties and thirties the beautiful work of Verney on water diuresis and the posterior pituitary (reviewed by Verney, 1957) directed attention to the control of extracellular tonicity and sodium concentration. In addition H. W. Smith, in his first great monograph on the kidney, published in 1937, suggested that volume control was merely a fortuitous consequence of the kidney's control of extracellular composition: 'The kidney itself is not concerned with whether two or five litres of plasma is circulating in the vascular bed, so long as that plasma has the proper composition. And this organ, in attempting to regulate the composition of the plasma, will frequently enlarge or reduce the volume of that fluid, and, indirectly, of the interstitial fluid, beyond physiological limits and sometimes to fatal excess.'

Not all were of this opinion, and both Peters (1935) and Borst (1938), writing at about this time, concluded that extracellular volume was under renal control, and suggested that the circulating blood volume was the factor somehow influencing renal activity. But Smith's book became the standard text in renal physiology and may have reduced interest in the subject of volume control.

In the last few years progress has advanced tremendously. Particular attention has been devoted to the metabolism of sodium, for a simple reason which stems directly from the work of Verney. In normal circumstances two finely adjusted

mechanisms—thirst, and the osmoreceptor-posterior pituitary apparatus described by Verney—so control the total amount of water in the body that the tonicity of extracellular fluid, and hence its concentration of sodium, remains constant. As long as these two mechanisms function in this way the volume of extracellular fluid will depend on its total content of sodium. The

[illegible]

out what controls excretion of sodium.

Sodium Excretion

The excretion of sodium does not lend itself to easy study. The rate varies considerably from hour to hour, and from day to day, being greatly affected by exertion, emotion, posture, and time of day, quite apart from the variations to be expected from changes in intake. When these variables are controlled as far as possible a fairly constant pattern appears after a few days, an example of which is shown in Figure 1. Usually the rate of excretion shows a marked daily rhythm, being greatest about midday and least at midnight, but the magnitude of this rhythm, the cause of which is quite obscure, varies greatly from person to person, and a few people show no rhythm at all.

Manœuvres designed to contract or expand extracellular fluid are followed by reciprocal changes in the excretion of sodium. These changes are sluggish, taking hours, days, or even weeks for completion, depending on the size of the original disturbance. Figure 2 shows an experiment in which the sodium diuresis following infusion of a litre of isotonic saline was virtually complete within 36 hours. By contrast Gamble (1951) has reported that a subject took two weeks to come back into sodium balance after increasing his salt intake by 5 grammes daily.

Redistribution of extracellular fluid within the body, without change in total volume, also markedly influences excretion of sodium. Generally speaking, steps which reduce the central intravascular volume, blood-pressure, or cardiac output, such as passive standing or obstruction to the venous return from the

in the field have given their attention to the subject. Claude Bernard, who originated our concept of extracellular fluid as an entity, specifically mentioned control of volume: 'Chez les animaux à vie libre il doit exister un ensemble de dispositions réglant les pertes et les apports de manière à maintenir la quantité d'eau nécessaire dans le milieu intérieur.' (Bernard, 1878.)

Thirty years later Starling (1909) recognized that intestinal absorption of salt and water proceeds regardless of the needs of the body, and pointed out that control of extracellular fluid volume must therefore devolve on the kidneys. But in spite of these auspicious beginnings the early part of the century saw an almost complete lack of speculation and experiment on this important subject. Instead, in the twenties and thirties the beautiful work of Verney on water diuresis and the posterior pituitary (reviewed by Verney, 1957) directed attention to the control of extracellular tonicity and sodium concentration. In addition H. W. Smith, in his first great monograph on the kidney, published in 1937, suggested that volume control was merely a fortuitous consequence of the kidney's control of extracellular composition: 'The kidney itself is not concerned with whether two or five litres of plasma is circulating in the vascular bed, so long as that plasma has the proper composition. And this organ, in attempting to regulate the composition of the plasma, will frequently enlarge or reduce the volume of that fluid, and, indirectly, of the interstitial fluid, beyond physiological limits and sometimes to fatal excess.'

Not all were of this opinion, and both Peters (1935) and Borst (1938), writing at about this time, concluded that extracellular volume was under renal control, and suggested that the circulating blood volume was the factor somehow influencing renal activity. But Smith's book became the standard text in renal physiology and may have reduced interest in the subject of volume control.

In the last few years progress has advanced tremendously. Particular attention has been devoted to the metabolism of sodium, for a simple reason which stems directly from the work of Verney. In normal circumstances two finely adjusted

by administration of antidiuretic hormone, a state of water intoxication develops which is characterized by expansion of both extracellular and intracellular fluids, hyponatraemia as a result of dilution, and marked increase in the excretion of sodium

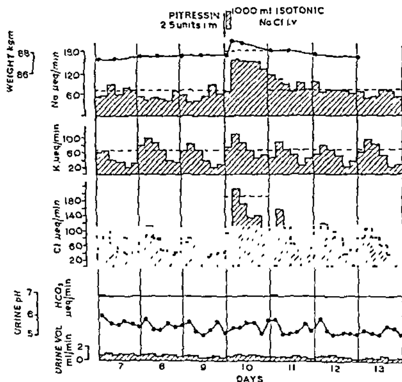


FIG. 2. Effect of an infusion of physiological saline on the excretion of electrolytes Pitressin tannate was given to prevent the water diuresis which has occasionally been reported following such an infusion, and to serve as a control with subsequent experiments. The dotted lines show the electrolyte intake averaged over each 24-hour period.

(Leaf, Bartter, Santos and Wrong, 1953). This sodium diuresis, of which examples are shown in Figures 3 and 4, will further lower the already subnormal serum sodium concentration, but is a logical consequence of the increase in extracellular volume.

Sodium is excreted by a process of glomerular filtration and

legs, also reduce sodium excretion (reviewed by Robinson, 1954). This suggests that the structures which are sensitive to changes in extracellular volume, somehow initiating parallel changes in sodium excretion, are really sensitive to some

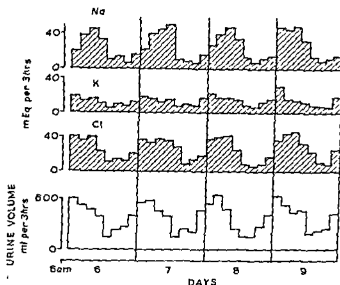


FIG. 1. Excretion of electrolytes and water by a young man at complete

property of central intravascular volume, a point we will return to later.

It is worth noting that the changes in sodium excretion which follow alterations in extracellular volume may sometimes exaggerate, rather than correct, abnormalities in serum sodium concentration. Thus pure water deprivation, if sufficiently prolonged, will lead to contraction of the extracellular space and an increase in serum sodium concentration—yet sodium disappears from the urine (Elkinton and Taffel, 1942). This renal response makes sense in terms of volume control, but is meaningless if we think only in terms of control of composition. Similarly when water is given in excess, and its excretion prevented

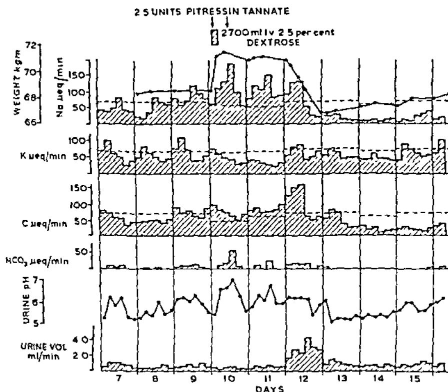


FIG. 4 Effect of pitressin-induced water retention on the excretion of electrolytes (a different subject) In this experiment hydration was maintained for 48 hours Note once more the delay between hydration and maximal diuresis of sodium and bicarbonate, and the close relationship between sodium excretion and body weight (Wrong, 1956)

changes of under 5 per cent in either filtered or reabsorbed sodium. Unfortunately our clearances techniques are not sufficiently accurate to detect changes of this order, and consequently there has been great uncertainty as to whether normal variations in the excretion of sodium are determined mainly by alterations in glomerular or tubular function.

Much of the earlier work on sodium excretion was done on the dog, mainly by Smith and his colleagues (reviewed by Smith, 1951). Now the dog has a very labile glomerular filtra-

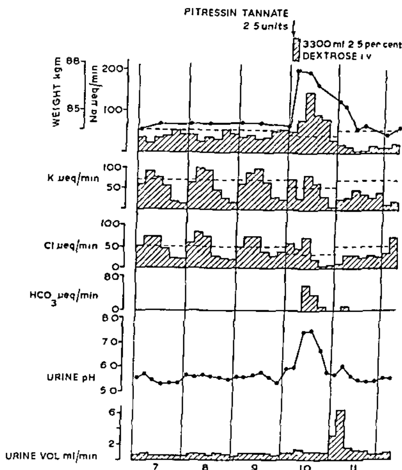


FIG 3 Effect of pitressin-induced water retention on the excretion of electrolytes Note (1) the delay of 8–12 hours between hydration and maximal sodium diuresis, (2) the simultaneous bicarbonate diuresis and increase in urinary pH (Reproduced from Wrong, 1956)

tubular reabsorption. Approximately 600 grammes, or seven times the total amount in the body, are filtered daily, but over 95 per cent of this vast amount is reabsorbed by the renal tubules. A moment's calculation will show that the whole range of sodium excretion of which the human kidneys are known to be capable (say from one half to fifteen hundred milliequivalents daily) could theoretically be brought about by

with or even exceeds intake (August, Nelson and Thorn, 1958). So although patients with cardiac, cirrhotic, and nephrotic oedema excrete excessive amounts of aldosterone in their urine, apparently something more than just increased aldosterone production is necessary for the massive sodium retention that gives rise to oedema. It is even possible that oedema could develop in the absence of aldosterone, for treatment of the oedema of hepatic cirrhosis with amphenone (which blocks adrenal synthesis of the hormone) or with steroids which antagonize the action of aldosterone on the kidney, has not always caused significant sodium diuresis.

Aldosterone labelled with tritium has a half-life in the circulation of only twenty minutes or so (Ayres *et al.*, 1958), but when injected intravenously it reduces sodium excretion for six to eight hours (Muller, Mach and Naegeli, 1955; Ross, Reddy, Rivera and Thorn, 1959). If injected slowly into a renal artery of the dog it has no effect on electrolyte excretion for between five and sixty minutes, and a maximum effect is seen after two to three hours. These observations suggest that alterations in aldosterone secretion are not responsible for rapid changes in sodium excretion such as take place after haemorrhage or change in posture. But changes in the rate of aldosterone secretion are fairly rapid for a humoral mechanism—the concentration of hormone in adrenal venous blood from the dog alters within thirty to forty-five minutes of haemorrhage or transfusion, and my Manchester colleagues, Drs. Gowenlock, Mills and Thomas (1959), have found that subjects excrete an increased amount of aldosterone in the urine within three hours of adopting a standing position. This last effect may be due to venous pooling in the lower part of the body, for it can be prevented by standing up to the neck in water, an entertaining class experiment.

The adrenal secretion of aldosterone is regulated by the state of body sodium and potassium, and is little if at all under anterior pituitary control. Deprivation of sodium, or potassium excess, increases the rate of secretion, whereas a potassium deficiency or sodium excess has the opposite effect. Alterations in potassium balance have less effect than comparable changes in sodium balance, and it has been suggested that any effect they

tion rate, which can be markedly increased by expansion of the extracellular space. This increased filtration is largely responsible for the sodium diuresis which follows an infusion of saline, although a reduction in tubular reabsorption also seems to occur. In man the filtration rate is more constant, and two groups of workers (Crawford and Ludemann, 1951; Green, Bridges, Johnson, Lehman, Gray and Field, 1950) have failed to find a significant increase during the sodium diuresis which follows intravenous infusion of isotonic sodium chloride. (Perhaps it is because of this fixity of filtration rate that man excretes a sodium load much more slowly than does the dog.) Black, Platt and Stanbury (1950) have shown in man that a definite increase in tubular reabsorption occurs after a few days on a low salt diet. But, as stated earlier, the accuracy of clearance techniques is not such that we can deny that alterations in filtration rate might appreciably contribute to normal fluctuations in sodium excretion.

Aldosterone

When aldosterone was isolated in 1953 it seemed the answer to many problems—a potent salt-retaining hormone which might account for all variations in sodium excretion. Since then we have learned that although there is much that aldosterone explains, there is also much that it fails to explain Some excellent reviews have recently appeared on this subject, so I will confine my remarks to a few points.

Aldosterone is much the most powerful naturally occurring mineralocorticoid so far discovered, but it is not the only one, and others have been shown to have importance in disease states and may play a part in normal homeostasis. In subjects on an average diet the adrenals secrete about 200 micrograms of aldosterone daily (Ayres, Barlow, Garrod, Kellie, Tait, Tait and Walker, 1958). When doses of this order are given to normal subjects or patients with Addison's disease a definite retention of sodium and loss of potassium follow. But it has proved impossible to cause oedema by administration of many times this amount of aldosterone, two to six hundred milliequivalents of sodium are retained and then sodium excretion catches up

possible that the hormone is produced by a centre here or that further intermediate steps exist. This work is all very exciting, but much of it is very recent and needs confirmation.

The suggestion has been made that aldosterone plays little

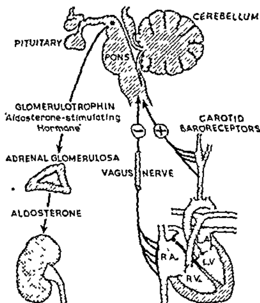


FIG. 5 Control of aldosterone secretion. For description see text.

or no part in day-to-day electrolyte balance. I do not believe that this is so, for a number of reasons:

(1) Addisonian patients who are treated with glucocorticoids alone often suffer from hypotension and hyponatraemia unless given extra salt. These abnormalities can be corrected by aldosterone in amounts similar to those secreted by the normal adrenal.

(2) Experiences common to all of us—changes in posture, or in dietary intake of sodium and potassium—cause alterations in the excretion of aldosterone.

(3) Urinary excretion of aldosterone correlates fairly well with the excretion of sodium and potassium, particularly with the potassium/sodium ratio (see Fig. 6). This suggests that

do have is mediated by the changes in sodium balance they bring about; however, work by Bartter (1956) suggests that this is not the whole explanation and there is some evidence suggesting that the serum concentration of potassium directly influences adrenal activity. With sodium we are on more certain ground, for Bartter, Liddle, Duncan, Barber and Delea (1956) have shown that this influences production of aldosterone through its effect on extracellular fluid volume. In fact, if extracellular volume is altered without any change in external sodium balance, as in a pure water deficit, or water intoxication, appropriate changes in aldosterone excretion ensue. The sodium diuresis of pitressin-induced water retention, for example, is associated with a fall in aldosterone excretion, and the delay of several hours before this diuresis becomes maximal (Figs. 3 and 4) may be due to the action of previously formed aldosterone. But normally extracellular volume follows sodium balance and aldosterone excretion varies in the opposite direction. In a series of ingenious experiments Bartter and others (1956) have shown that the intravascular portion of extracellular fluid is the critical component affecting aldosterone production.

Figure 5 shows some of the central pathways which may be concerned in the control of aldosterone secretion, as elucidated from animal experiments, chiefly in the dog. Bartter, Mills and Gann (1959) have produced evidence that impulses leading to *increased* secretion are transmitted from baroreceptors situated on the main branches from the thoracic aorta. On the other hand Farrell and his colleagues (Anderson, McCally and Farrell, 1959) have shown that pulling on the right auricle causes a *reduction* in the rate of secretion, presumably through stimulation of atrial stretch receptors, and observations of Mills, Casper, and Bartter (1958) suggest that afferent impulses for this effect travel in the vagus nerve. A number of separate workers (reviewed by Farrell, 1958) have produced evidence that aldosterone secretion by the adrenal glomerulosa is under hormonal control, and the hormone has been tentatively named 'glomerulotropin' or 'aldosterone-stimulating hormone'. Farrell has shown that production of this substance ceases after destruction of a small area in the posterior hypothalamus, and it is

possible that the hormone is produced by a centre here or that further intermediate steps exist. This work is all very exciting, but much of it is very recent and needs confirmation.

The suggestion has been made that aldosterone plays little

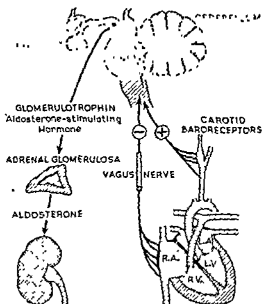


FIG. 5. Control of aldosterone secretion For description see text

or no part in day-to-day electrolyte balance. I do not believe that this is so, for a number of reasons:

(1) Addisonian patients who are treated with glucocorticoids alone often suffer from hypotension and hyponatraemia unless given extra salt. These abnormalities can be corrected by aldosterone in amounts similar to those secreted by the normal adrenal.

(2) Experiences common to all of us—changes in posture, or in dietary intake of sodium and potassium—cause alterations in the excretion of aldosterone.

(3) Urinary excretion of aldosterone correlates fairly well with the excretion of sodium and potassium, particularly with the potassium/sodium ratio (see Fig. 6). This suggests that

aldosterone is at least one of the factors controlling normal electrolyte excretion.

I have been particularly interested in the possible role of aldosterone in the excretion of hydrogen ion. Aldosterone probably acts on the site in the renal tubule where sodium is re-

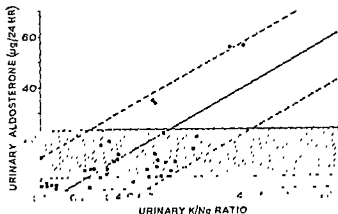


Fig. 6. Relationship between the urinary aldosterone of normal subjects, and urinary K/Na. Unpublished observations of Dr. A. H. Gowenlock. Urinary aldosterone was determined by method of Ayres, Garrod, Simpson and Tait (1957); the shaded area shows their normal range for 31 male subjects taking average diet.

absorbed in exchange for potassium and hydrogen ions; it might therefore be expected to cause increased excretion of hydrogen ion (or retention of bicarbonate ion, which physiologically comes to the same thing) at the same time as increased tubular absorption of sodium. Extracellular fluid contains about 26 m.-equiv. per litre of bicarbonate, and if aldosterone increased the excretion of hydrogen ion it would tend to preserve the composition of extracellular fluid while promoting its expansion. Bartter has shown, in fact, that aldosterone does acutely increase hydrogen ion excretion. Suppression of endogenous aldosterone production might therefore be expected to lead to a bicarbonate diuresis, and this might be the explanation for the increased excretion of bicarbonate which accompanied sodium in the experiments shown in Figures 3 and 4. A more marked bicarbonate diuresis is shown in Figure 7 to accompany

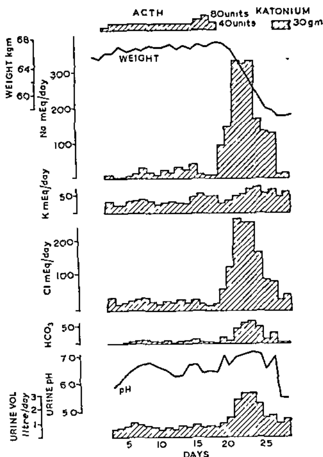


FIG. 7 Nephrotic syndrome, diuresis induced by corticotrophin The subject was a woman aged 31, receiving a low salt diet. Coincident with a 10 Kg weight loss there was a diuresis of sodium, chloride and bicarbonate in proportions corresponding to extracellular fluid

the sodium diuresis of a nephrotic patient who was losing her oedema following administration of corticotrophin. There is evidence that in this situation the diuresis, of what is essentially extracellular fluid, is due to suppression of endogenous aldosterone (Luetscher, Deming and Johnson, 1952).

Aldosterone also seems to effect the form in which hydrogen

ion is excreted, increasing the amount which appears in the urine combined with ammonia as ammonium ion, and reducing the amount combined with urinary buffer as titratable acid. Patients with 'primary aldosteronism' usually excrete a relatively alkaline urine containing large quantities of ammonium, and this abnormality has been attributed to tubular damage resulting from potassium deficiency. Dr. Gowenlock and myself have studied similar patients in Manchester and found that the abnormality reverted to normal as soon as the source of excess aldosterone was removed. This happened before any potassium was retained, which suggests that aldosterone itself, and not the accompanying potassium deficiency, was responsible.

Sodium Excretion—Other Influences

Recently a group of workers in Ann Arbor (Vander, Malvin, Wilde and Sullivan, 1958) have put forward an important new concept; it has not yet had time to receive the inevitable drubbing from other physiologists, and consequently some of its fallacies may have escaped me. Using their 'stop-flow' technique (which I have no room to describe here) these workers came to the conclusion that proximal tubular reabsorption of sodium is not an active metabolic process, as generally thought on rather inconclusive grounds, but a passive shift of isotonic fluid produced by the oncotic pressure of plasma protein in the peritubular capillaries. This is no more than the theory that Starling (1896) put forward so successfully over sixty years ago to explain equilibrium between plasma and interstitial fluids across the capillary membrane—at the arteriolar end of the capillary (that is, in this case, the glomerulus) hydrostatic pressure forces filtration of a protein-free fluid, but at the venous [illegible] ry) fluid [illegible] plasma [illegible] le when we remember that the glomerulus starts off like any other tuft of capillaries, and the primitive vertebrate renal tubule, both embryologically and phylogenetically, is in continuity with the coelomic cavity, part of the general interstitial space. The Ann Arbor workers point out that the oncotic pressure in the peri-

tubular capillaries depends largely on the extent to which the plasma proteins have been concentrated in the process of glomerular filtration. When a large part of the total renal plasma flow is filtered (that is, when the filtration fraction is high) proximal tubular reabsorption of sodium will be more complete than when the filtration fraction is low. Now it is well established that in chronic congestive heart failure the plasma flow is usually reduced to a greater extent than the filtration rate, with the result that the filtration fraction is high. This the authors believe to be responsible for the avid renal conservation of sodium seen in heart failure.

This theory is attractive, and one wonders why no one has seriously considered it before. Vander and others have not claimed that this mechanism plays any part in normal variations

could account for the sodium retention of the nephrotic syndrome, for here plasma oncotic pressure is well below normal. The key question is—How in cardiac failure is the glomerular filtration rate maintained (or only slightly reduced) despite marked reduction in renal plasma flow?

Dissociation between Sodium and Water Balance

So far, in discussing the control of extracellular volume, I have assumed that thirst and the posterior pituitary are unaffected by changes in volume, and function at all times so as to maintain the usual proportion of water to solute in extracellular fluid—control of extracellular volume then depends on control of sodium excretion. This is not entirely true, for contraction of extracellular volume, without any change in composition, may stimulate both thirst and renal conservation of water, and these in turn will tend to restore volume at the expense of osmolality. It is classical teaching, for instance, that severe haemorrhage results in thirst, it also causes a marked reduction in water excretion. McCance's classic study of experimental salt depletion, reported over twenty years ago (McCance, 1936), showed much the same thing. McCance found that subjects who were

progressively depleted of sodium by sweating, lost water (shown by their loss of weight) in proportion to their loss of sodium until the sodium deficit was about 350 m.-equiv. Thereafter their loss of water lagged behind that of sodium, permitting their serum sodium concentration to fall, and they lost their ability to have a water diuresis. The evidence from these and similar experiments (well reviewed by Strauss, 1957) suggests that contraction of extracellular volume may stimulate the posterior pituitary, but unfortunately there is no really satisfactory way of measuring the secretion of antidiuretic hormone. Recent work by Henry and Pearce (1956) has pointed to one mechanism by which this effect might be accomplished. They demonstrated the existence of stretch-receptors in the left atrium of the dog and cat which are capable of causing a water diuresis; it is likely that this is mediated through the posterior pituitary. Note that the suggestion here is of a left atrial receptor influencing excretion of *water*, whereas the receptor suggested by Farrell is in the right atrium and influences excretion of *sodium* through its effect on aldosterone secretion.

Conclusion

The control of extracellular volume may seem an unnecessarily complicated subject. But simplicity of design seems to have no particular biological advantages, and it might be taken as a general rule that the most vital functions, of which extracellular volume is surely one, are not controlled by a single mechanism. Certainly the subject will become even more complicated, but will probably turn out to be no more so than the control of many other vital processes.

REFERENCES

Anderson, C H , McCally, M. and Farrell, G. L. (1959). *Endocrinology*, 64, 202.

August, J. T., Nelson, D. H. and Thorn, G W. (1958). *J. clin. Invest* 37, 1549

Ayres, P. J., Barlow, J., Garrod, O , Kellie, A. E , Tait, S. A. S , Tait, J. F. and Walker, G. (1958). In *International Symposium on Aldosterone*. Churchill, London.

Ayres, P. J , Garrod, O., Simpson, S. A. and Tait, J. F. (1957). *Biochem. J.* 65, 639

Bartter, F. C (1956). *Metabolism*, 5, 369.

Bartter, F. C , Liddle, G W , Duncan, L. E., Barber, J. K. and Delea, C (1956). *J. clin. Invest* 35, 1306

Bartter, F. C , Mills, I. H. and Gann, D G (1959) *J. clin. Invest* 38, 986.

Bernard, C (1878) *Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux*. Baillière, Paris.

Black, D A. K , Platt, R. and Stanbury, S. W (1950) *Clin. Sc.* 9, 205.

Borst, J. G. G. (1938). *Acta med. scand.* 97, 68.

Conway, E. J. and McCormack, J. I. (1953). *J Physiol.* 120, 1.

Cooke, R. E., Segar, W. E., Cheek, D. B , Coville, F. E. and Darrow, D C (1952). *J. clin. Invest.* 31, 798

Cort, J. H. and [illegible]

[illegible], 1456.

[illegible] 266.

[illegible] *Amer. J. Physiol*

[illegible]

Edelman, I S , Leibman, J , Omeara, M. P. and Birkenfeld, L. W (1958). *J. clin Invest.* 37, 1236

Elkinton, J. R and Taffel, M (1942) *J clin Invest.* 21, 787.

Farrell, G. L (1958). *Physiol Rev.* 38, 709.

Gamble, J. L. (1951) *Stanf. Univ Publ , Univ Ser. Med Sci* 5, no. 1.

Gowenlock, A H , Mills, J N and Thomas, S (1959). *J. Physiol.* 146, 133

Green, D. M , Bridges, W. C , Johnson, A. D., Lehman, J. H , Gray, F. and Field, L. (1950) *Amer. J Physiol.* 160, 306

Henry, J. P and Pearce, J W (1956). *J. Physiol* 131, 572.

Hodgkin, A L , and Keynes, R. D (1955) *J Physiol.* 128, 28.

Laragh, J H (1954). *J. clin. Invest.* 33, 807

Leaf, A. (1956) *Biochem J* 62, 241

Leaf, A , Anderson, J. and Page, L B. (1958). *J gen. Physiol* 41, 657

Leaf, A , Bartter, F C , Santos, R. F and Wrong, O (1953). *J. clin. Invest* 32, 868.

Leaf, A , Chatillon, J Y , Wrong, O and Tuttle, E P (1954). *J. clin. Invest* 33, 1261

LUETSCHER, J. A., DEMING, Q. B. and JOHNSON, B. B. (1952). *In Ciba Foundation Colloquia on Endocrinology*, vol. iv. Churchill, London.
MCCANCE, R. A (1936). *Lancet*, i, 823.
MAFFLY, L. H. and LEAF, A. (1958). *Nature*, 182, 60.
[illegible]
20, 113.
[illegible]
ROSS, E. J., REDDY, W. J., RIVERA, A. and THORN, G. W. (1959). *J. clin. Endocrinol.* 19, 289.
SABBATANI, L. (1901). *J. physiol et pathol. gén.* 3, 939.
SMITH, H. W. (1932). *The Physiology of the Kidney.* Oxford University Press.
SMITH, H. W. (1951). *The Kidney.* Oxford University Press.
STARLING, E. H. (1896). *Lancet*, i, 1267.
STARLING, E. H. (1909). *The Fluids of the Body.* Constable, London
STRAUSS, M. B (1957). *Body Water in Man.* Churchill, London.
VANDER, A J., MALVIN, R. L, WILDE, W. S. and SULLIVAN, L. P. (1958). *Amer. J. Med.* 25, 497.
[illegible]

XXII

Biochemical Studies in Hepatic Coma

J. M. WALSHE

In the words of Ortega y Gasset we must 'arrive at the new truth with hands blood-stained from the slaughter of a thousand platitudes'. Now while I think it is likely that I shall be able to expose a few fallacies in the enormous literature on the subject of hepatic coma that must be covered in this lecture I doubt if I will be able to slay sufficient numbers of platitudes to qualify for the discovery of a new truth.

Before we can hope to understand the biochemical lesion underlying hepatic coma it is necessary to have a clear picture of the normal metabolic pathways involved. It is now generally agreed that, despite the structural changes found in the brains of patients dying in hepatic coma, the essential lesion causing the disturbance of consciousness is biochemical. Furthermore, it has been established that the syndrome of hepatic coma can occur in the presence of normal, or relatively normal, liver function In the majority of cases the primary lesion is in the hepatic parenchymal cells leading to a failure of the metabolic or homeostatic functions of the liver. In a few cases the defect appears to be in the hepatic circulation resulting in a shunting of the portal blood into collateral channels. In other words coma may be due either to a failure of the detoxifying functions of the liver or of its ability to add to the blood essential metabolites necessary to maintain normal cerebral functions.

Cerebral Metabolism

Under normal conditions the brain obtains its energy by the oxidation of glucose first by anaerobic glycolysis to pyruvate

and then via the tricarboxylic acid cycle to carbon dioxide and water. By this process it obtains the energy necessary for the synthesis of acetyl choline, for the phosphorylation of glucose, for the synthesis of creatine phosphate and adenosine triphosphate (ATP), for the incorporation of radiophosphorus into phospholipids and other organic phosphorus compounds, for the synthesis of glutamine for the maintenance of the ionic gradients across the cell membranes and also to sustain the electrical excitability of the neurones and the axons. It has been shown, both *in vitro* and *in vivo*, that the respiratory quotient of the brain approaches unity and it has also been shown *in vivo* that the cerebral uptake of oxygen corresponds to the utilization of glucose. There is no evidence that the brain obtains energy by the oxidation of fatty acids.

The normal human brain has a blood flow of 54 ml. per 100 g. *per minute, an oxygen uptake of 3·3 ml./100 g./min.*, an arteriovenous (A-V) oxygen difference of 6·3 volumes per cent and respiratory quotient of 0·99 according to Kety and Schmidt (1948). From these figures we can calculate the metabolic quotient of the brain and this comes to approximately 12 µl. oxygen per mg. dry weight brain per hour, a very similar figure to that obtained by McIlwain (1953) in his *in vitro* studies with human cortex slices. Scheinberg and Stead (1949) found slightly higher figures for oxygen uptake, 2·84 ml./100 g./min. and a glucose utilization of 6·27 mg./100 g./min., that is 81 mg. for a whole brain per minute. It has been shown by *Mangold, Sokoloff, Conner, Kleinermann, Therman* and Kety (1955) that there is no significant change in oxygen uptake during sleep, and likewise that there is no increase in cerebral oxygen utilization during mental activity (Sokoloff, Mangold, Wechsler, Kennedy and Kety, 1955).

Now although glucose is the normal fuel for the brain it has been shown *in vitro* that other sugars will support metabolic activity equally well, as also will pyruvate and lactate. It is surprising therefore to learn that in the hepatectomized dog only glucose was able to support electrical excitability of the cerebral cortex or to correct hypoglycaemic coma. Fructose, galactose and hexose diphosphate were inactive as also were

pyruvate, lactate, succinate and malonate (Maddocks, Hawkins and Holmes, 1939; Elliott and Henry, 1946). *In vitro* these compounds can supply a comparable number of high-energy phosphate bonds to glucose. Their failure to correct hypoglycaemic symptoms has been attributed by Himwich (1951) to an inability to cross the blood brain barrier, but Coxon (1952) has suggested that the capacity of glucose to accept phosphate may be important. Certainly ATP is necessary for this step and its breakdown to ADP and inorganic phosphate will in turn activate the enzyme glutaminase with the resultant formation of ketoglutarate which enters the Krebs cycle and supplies further energy for the resynthesis of ATP.

It is this process of substrate oxidation, or more precisely electron transfer, that supplies the energy of the cell. The transfer of two hydrogen atoms to oxygen yields three and possibly four high-energy bonds. The glycolytic breakdown of one molecule of glucose to two of pyruvate gives eight such bonds and a further fifteen result from the complete oxidation of each molecule of pyruvate via the Krebs cycle. That is 38 high-energy phosphate bonds are formed during the complete oxidation of glucose. Assuming 12,000 calories per bond, then 465,000 calories are stored as phosphate bonds for each gram molecule of glucose oxidized, or 67 per cent of the calculated free energy loss involved (Strickland, 1956). Obviously any slowing down of glucose oxidation via the Krebs cycle will seriously reduce the energy available to the cells

One of the more vulnerable points in the cycle appears to be that at which pyruvate condenses with oxaloacetate to form [illegible]

pyruvate and then transfers the acetyl group to coenzyme A, resulting in the formation of acetyl co-A and reduced lipothiamide pyrophosphate. The acetyl group condenses with oxaloacetate to form citrate with the restoration of the reduced

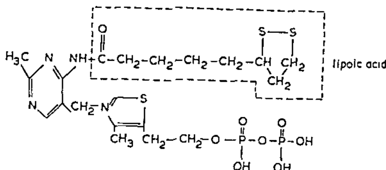


Fig. 1. Lipothiamide pyrophosphate (LTPP).

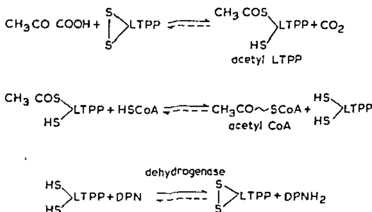


Fig 2. Postulated scheme for the formation of Acetyl CoA. (After Reed and Debusk.)

coenzyme A. In the meantime the lipothiamide cedes hydrogen to DPN and is thus able to repair its disulphide bond (Fig. 2). A similar mechanism is probably operative for the oxidation of ketoglutarate to succinyl CO-A and succinic acid.

Ammonia Binding Mechanism

Now although cerebral energetics are centred largely around carbohydrate metabolism there are other features of cerebral metabolism which must be considered. Of these perhaps the most important is the role of glutamic acid in the ammonia

binding mechanism. The normal brain contains ammonia and the exact amount varies with the method by which the animal has been killed. The lowest figures reported are those of Richter and Dawson (1948) who showed that after barbiturate anaesthesia, if the animal was dropped into liquid air, the average brain ammonia was 60 μg./100 g. brain. After electrically or chemically produced convulsions the concentration could be as high as 800 to 1,000 μg. Figures for animals killed by decapitation, and previously in a normal state of activity, are of the order of 200 to 400 μg. per cent in most series (Benitez, Pscheidt and Stone, 1954; Richter and Dawson, 1948). *In vitro* studies have also shown that the brain will make ammonia and that glucose is necessary for the removal of ammonia so formed. Weil Malherbe and Green (1955) recently studied the factors involved and showed that the amide nitrogen of glutamine was not the source of the ammonia, even in the absence of glucose. The addition of high-energy phosphate or diphophopyridine nucleotide (DPN) did not influence ammonia synthesis nor could they link it with any of the five known brain deaminating enzymes. They concluded that the formation of ammonia was due to the action of a very unstable proteinase present in brain. They also reported the failure of fluoroacetate to increase the ammonia formation by brain slices, although it has this action in the intact animal (Benitez *et al.*, 1954). Indeed it has been postulated that the mechanism of fluoroacetate toxicity might be due to the accumulation of ammonia in the cell. This occurs before the onset of convulsions, or before the accumulation of citrate, which in fact it exceeds on a molar basis.

The normal mechanism for the removal of ammonia formed in the brain is believed to be by combination with glutamate to form glutamine; this process requires high-energy phosphate. There is much evidence to support this theory, postulated by Krebs in 1935. There is a high concentration of glutamic acid in the brain, in most species around 100 mg. per 100 g. wet weight, while glutamine is usually present in about half the concentration. Given glucose as substrate the brain will form glutamine while, in the absence of substrate, ammonia accumulates. On the other hand both Richter and Dawson (1948) and

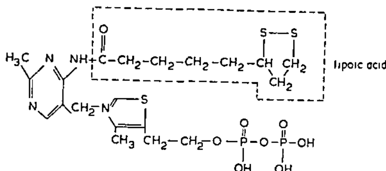
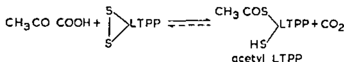
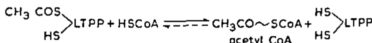


FIG. 1. Lipothiamide pyrophosphate (LTPP).





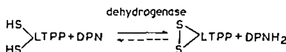


FIG 2 Postulated scheme for the formation of Acetyl CoA (After Reed and Debusk)

coenzyme A. In the meantime the lipothiamide cedes hydrogen to DPN and is thus able to repair its disulphide bond (Fig. 2). A similar mechanism is probably operative for the oxidation of ketoglutarate to succinyl co-A and succinic acid.

AMMONIA BINDING MECHANISM

Now although cerebral energetics are centred largely around carbohydrate metabolism there are other features of cerebral metabolism which must be considered Of these perhaps the most important is the role of glutamic acid in the ammonia

pathways is, I believe, necessary for an understanding of what is to follow.

METABOLIC LESION IN HEPATIC COMA

In the last decade the literature on hepatic coma has multiplied in a remarkable fashion and from it certain facts have crystallized. In the Continental literature of the 1930s (Fuld, 1933; Van Caulaert, Deviller and Halff, 1932a, b; Van Caulaert, Deviller and Hofstein, 1932; Kirk, 1936) attention was drawn to the raised concentration of ammonia that occurred in the blood of patients with liver disease and also the mental symptoms which could be induced in susceptible patients by the administration of ammonium salts. More recently the work of Davidson and his team in Boston (Webster and Davidson, 1957; Seegmiller, Schwartz and Davidson, 1954) and Sherlock in this country has clarified the picture and gone a long way towards defining the salient facts, although our interpretation of them is still not certain. Shortly we shall have to consider the place of ammonia in the genesis of hepatic coma and to review the evidence for and against its role as a cerebral intoxicant. Another compound which has been found to induce coma in susceptible subjects is methionine. Later we shall also have to consider what role susceptibility plays in the development of hepatic coma. Other factors have also been incriminated; these are the so-called toxic amines formed by bacterial decarboxylation of aminoacids in the gut, amines which in turn might give rise to the more toxic aldehydes under the influence of amine oxidase. Coproporphyrin metabolism is also disturbed in hepatic disease, and in severely jaundiced patients the concentra-

lites has been less extensively studied (Geiger, Dobkin and Magnes, 1953). Geiger and Yamasaki (1956) have shown that the electrical excitability of the perfused brain depends upon the presence of the liver in the circulation or on the infusion of the nucleosides cytidine and uridine. This suggests that these are normally supplied by the liver and are essential for cerebral metabolism.

Benitez and others (1954) found no increase in cerebral glutamine after convulsions or after the injection of ammonium salts. Again, although glutamic acid can enter the nerve cell *in vitro*, it is almost certain it cannot do so in the intact animal so that the glutamic acid present in the brain must be formed either from glutamine, which can cross the blood brain barrier readily, or from ketoglutarate; the enzyme glutamic dehydrogenase can carry through this latter reaction without energy, in the presence of excess ammonia. Both glutamate and glutamine are active in transamination reactions. With pyruvate, glutamate can form ketoglutarate and alanine and with oxaloacetate it will form ketoglutarate and aspartic acid. Moreover two molecules of glutamate can oxidize to form one of glutamine and one of ketoglutarate. Thus glutamate can act as an acceptor or a donor of amino groups according to the needs of the cell and it can also supply ketoglutarate for the Krebs cycle. *In vitro* both glutamate and glutamine can act as substrates to support cerebral oxidation and it is curious therefore that when no substrate is supplied there is only a small fall in the glutamate concentration of the cell and no fall in the glutamine level. A still more remarkable fact is that under conditions when ammonia increases rapidly in the cell the glutamine concentration does not show a corresponding rise as has been noted by Waelsch (1951), by Richter and Dawson (1948) and by Benitez and others (1954). In addition to its role in transamination and amidation, glutamate also undergoes decarboxylation with the formation of gamma aminobutyric acid; this is believed to be an intracellular buffering mechanism leading to the freeing of fixed base to neutralize organic acids formed during active metabolism. In the human cortex McIlwain (1953) has shown that glutamate, in the absence of glucose, can support respiration and the response to electric excitation, a function it could not maintain in a rat, guinea-pig or rabbit brain. By supplying amino groups to large numbers of ketoacids and by amidation to glutamine, glutamate also plays a key role in protein syn-

pathways is, I believe, necessary for an understanding of what is to follow.

Metabolic Lesion in Hepatic Coma

In the last decade the literature on hepatic coma has multiplied in a remarkable fashion and from it certain facts have crystallized. In the Continental literature of the 1930s (Fuld, 1933; Van Caulaert, Deviller and Halff, 1932a, b; Van Caulaert, Deviller and Hofstein, 1932; Kirk, 1936) attention was drawn to the raised concentration of ammonia that occurred in the blood of patients with liver disease and also the mental symptoms which could be induced in susceptible patients by the administration of ammonium salts. More recently the work of

defining the salient facts, although our interpretation of them is still not certain. Shortly we shall have to consider the place of ammonia in the genesis of hepatic coma and to review the evidence for and against its role as a cerebral intoxicant. Another compound which has been found to induce coma in susceptible subjects is methionine. Later we shall also have to consider what role susceptibility plays in the development of hepatic coma. Other factors have also been incriminated; these are the so-called toxic amines formed by bacterial decarboxylation of aminoacids in the gut, amines which in turn might give rise to the more toxic aldehydes under the influence of amine oxidase. Coproporphyrin metabolism is also disturbed in hepatic disease, and in severely jaundiced patients the concentra-

lites has been less extensively studied (Geiger, Dobkin and Magnes, 1953). Geiger and Yamasaki (1956) have shown that the electrical excitability of the perfused brain depends upon the presence of the liver in the circulation or on the infusion of the nucleosides cytidine and uridine. This suggests that these are normally supplied by the liver and are essential for cerebral metabolism.

We have seen already that the normal brain receives a blood flow of 54 ml. per 100 g. and takes up oxygen at the rate of 3·3 ml. per 100 g. per minute. Fazekas, Ticktin, Ehrmantraut and Alman (1956) have shown that in hepatic coma the blood flow is reduced to 40 ml. a minute, a fall of 26 per cent, whereas the oxygen uptake is only 1·6 cc., a fall of 50 per cent below normal. In other words glucose utilization falls out of proportion to the diminished blood supply to the brain. However, as Fazekas and others (1956) pointed out, there was a large degree of overlap both for blood supply and oxygen utilization in the comatose and non-comatose patients they studied.

Ammonia as a Possible Factor

There is a similar overlap among comatose and non-comatose patients when they are grouped into those with a high and those with a normal blood ammonia and in most series about 10 per cent of patients in coma fall within the normal range (Sherlock, 1958; Webster and Gabruzda, 1958). However, Bessman and Bessman (1955) have pointed out that this poor correlation is due, at least in part, to using peripheral venous and not arterial blood for the ammonia determinations and they have shown that in most patients in hepatic coma the arterial ammonia level is elevated and that the tissues are taking up ammonia. The rate of uptake appears to be proportional to the degree of elevation of the arterial ammonia concentration (Webster and Gabduzda, 1958). In normal subjects and in patients with liver disease whose blood ammonia level is not raised there is an ammonia equilibrium across the brain although some brains may take up or evolve small amounts of this compound. On the other hand in patients in hepatic coma whose blood ammonia is high the brain as well as the peripheral tissues take up ammonia (Bessman and Bessman, 1955; Fazekas et al., 1956; Webster and Gabduzda, 1958; Summerskill, Wolfe and Davidson, 1957), an average figure for ammonia extraction being 50 μg./100 ml. blood. Taking Fazekas' figures for cerebral blood flow in comatose patients as 40 ml. a minute per 100 g brain, or approximately 500 ml. for an average human brain, it means that the cerebral uptake of ammonia is 250 μg. per minute.

Now one may ask if this uptake of ammonia by the peripheral tissues is peculiar to hepatic coma? The answer is definitely no. Tyor and Wilson (1958) have demonstrated that the blood ammonia concentration can be raised in normal subjects well into the coma range by the infusion of ammonium salts. They gave either ammonium chloride or lactate in amounts up to 94 m. eq. in one hour and achieved blood levels of over 700 µg. of ammonia nitrogen per 100 ml. without producing untoward symptoms. At these high concentrations the muscles removed ammonia in proportion to the degree of elevation and arterio-venous differences of over 300 µg. per 100 ml. were found, but the mechanism for removal of ammonia was readily saturated with high loads. Both Stauffer and Scribner (1957), and Wolfe, Fast, Stormont and Davidson (1958) have reported the production of a syndrome like hepatic coma in patients without liver disease, but in metabolic alkalosis, by the administration of ammonium chloride. Unfortunately, information as to the cerebral uptake of ammonia by these patients is not available. However, it is clear from details given in the latter paper that peripheral uptake of ammonia must have occurred. The patient had pyloric stenosis and persistent vomiting. He was given 29 g. of ammonium chloride in 20 hours, that is, 7·6 g. of ammonia nitrogen. Now if this were to be distributed in 10 litres of extracellular fluid (a generous allowance for a severely dehydrated man) it is equivalent to 76 mg. ammonia nitrogen per 100 ml. At the end of this infusion the blood ammonia was only 127 µg./100 ml., so that in 20 hours he had removed some 7·58 g. of ammonia nitrogen from his extracellular fluid, that is at the rate of approximately 3,790 µg./100 ml. per hour for 20 hours, no small feat when it is also claimed that specific therapy with arginine produced a further lowering of 70 µg./100 ml. in the twenty-first hour.

The mechanism of ammonia removal by the muscles could be either by incorporation into carbamyl phosphate, as a precursor of the Krebs urea cycle, by amidation of glutamate to form glutamine, or by the action of glutamic dehydrogenase to form glutamate from ketoglutarate. As the formation of carbamyl phosphate is inhibited by Diamox this might explain the

ability of Diamox to induce coma in susceptible subjects (Bessman, 1959). Similarly the formation of glutamate from ketoglutarate could reduce the availability of this ketoacid for the tricarboxylic acid cycle. That this is the toxic action of ammonia has been suggested by Bessman and Bessman (1955) but the key point here would seem to be whether the glutamate in the brain could hand on the unwanted amino group by way of the transaminase reactions already mentioned (Weil Malherbe, 1952). This would result in the restoration of ketoglutarate, which is probably not needed in more than catalytic amounts to keep the cycle operative.

Earlier we saw that the brain might remove an average of 250 µg. ammonia from the blood every minute. Now each microgram of ammonia nitrogen is equivalent to 10 µg. of glutamic acid. The brain contains 100 mg. of this aminoacid per 100 g., or 1·3 g. for an average adult brain, so that it would take over 9 hours at this rather high rate of ammonia extraction for the brain to deplete itself of glutamate assuming that no replacement mechanism was available. Weil Malherbe (1952) has pointed out the risk of dangerous depletion of ketoglutarate by overactivity of the ammonia binding mechanism but considered the ketoacid could be replaced by pyruvate or oxaloacetate accepting amino groups from glutamate. Bessman (1956), however, considers that this mechanism, while operative in muscle and liver, is not available to the brain. These contradictory views may well be due to species differences in the experimental animals studied since I have been able to show that glutamic-oxaloacetic transaminase and glutamic-pyruvic transaminase are present in the rat brain in concentrations comparable to those in liver.

This point wants settling for human brain. It would also be interesting to know the concentration of glutamate and ketoglutarate in the brains of patients dying in hepatic coma. At present there is no evidence of ketoacid depletion in such patients. Not only is the blood pyruvate elevated but so also is the ketoglutarate (Walshe, 1955; Dawson, De Groote, Rosenthal and Sherlock, 1957; Summerskill, Wolfe and Davidson, 1957). Dawson and others (1957) could find no evidence for

depletion of ketoglutarate after giving ammonium chloride to cirrhotic patients and Davidson's group showed that administration of ammonium salts actually resulted in an elevation of the blood ketoglutarate levels. They believed that the excess ammonia inhibited final glucose oxidation with the resultant accumulation of ketoacids. A more probable explanation is that,

possible mechanisms of ammonia toxicity have been suggested. Formation of glutamine requires ATP and abnormal activity of this system might lead to a deficiency for other systems requiring ATP (Weil Malherbe, 1948), such as acetyl choline synthesis (Mann, Tennenbaum and Quastel, 1939), incorporation of phosphorus into phospholipids (Findlay, Rossiter and Magee, 1954) and the maintenance of creatine phosphate levels and of electrical excitability (Gore and McIlwain, 1952). Ammonia also appears to inhibit the fixation of carbon dioxide by pyruvate to form oxaloacetate but Crane and Ball (1951), who investigated this, could find no evidence to suggest that ammonia combined with the ketoacids to form glutamate or alanine.

However, our understanding of the role of ammonia in the genesis of hepatic coma is further obscured by the observation, made originally by Webster (1956) and since confirmed by Summerskill and others (1957), that in some cases of hepatic coma the brain may actually evolve ammonia into the blood. It appears to be those patients whose coma has not resulted from a protein load who most often fall into this group.

Cerebral ammonia output as high as 50 μg. nitrogen per 100 g. brain per minute has been recorded by Webster and Gabduzda(1958) and no fewer than ten out of fourteen of their patients whose coma was spontaneous showed this phenomenon. On the other hand all of their patients in whom coma was precipitated by gastro-intestinal bleeding showed an uptake of ammonia by the brain. Summerskill and others (1957) have postulated that a progressive falling off of ammonia uptake occurs in severe hepatic disease resulting in a negligible uptake,

equilibrium and finally output of ammonia by peripheral tissues and brain, despite the high arterial level.

Probably the systems for removing ammonia which are normally active in liver and peripheral tissue break down in hepatic coma and the degree of their reserve conditions the susceptibility of the patient to an abnormal ammonia load. This could explain the poor clinical correlation often found between blood ammonia levels and the state of consciousness and the observation that the normal individual can tolerate ammonia concentrations far into the coma range.

Evolution of ammonia by the brain of patients in hepatic coma is in keeping with the *in vitro* studies of Weil Malherbe and Green (1955) who showed that the brain was capable of producing large quantities of ammonia. I have also studied this (Walshe, De Carli and Davidson, 1958b) and it is clear that ammonia is not an abnormal cerebral metabolite. The more favourable the *in vitro* conditions and the greater the rate of oxygen uptake, the less the amount of ammonia present in the cortex slice. This is presumably due to the optimum function of the ammonia binding mechanism. If the brain slice is deprived of glucose ammonia production rises, but if the brain is supplied with glutamine or glucosamine as substrate oxygen uptake increases while ammonia production is greatly augmented. Glucose and glutamine together support oxygen uptake well and also result in a very high rate of ammonia production equivalent to approximately 3 mg. per whole brain per minute. That is, the production of ammonia, in the cell, at its most vulnerable point, does not inhibit oxidation in any way (Walshe *et al.*, 1958). On the other hand glucosamine and glucose result in a high rate of ammonia production and also inhibit oxygen uptake, an observation previously made by Harpur and Quastel (1949) who believed the mechanism was a competition between glucose and glucosamine for phosphorylase. As this enzyme acts more slowly on glucosamine the ability of the brain to phosphorylate glucose is inhibited, hence reducing the supply of energy by the Krebs cycle. It is a curious point that the brain will apparently deaminate added glutamine in the presence of glucose but will not deaminate the endogenous compound,

suggesting that this is loosely bound in some way so that it is not readily available for the metabolic pool.

There is evidence that the mechanism for ammonia removal is different in cirrhotic patients from normals. Bessman, Shear and Fitzgerald (1957) have shown that neither arginine nor glutamate affected the rate of removal of infused ammonia from the blood of non-cirrhotic patients. But in patients with cirrhosis arginine slightly increased the rate of ammonia removal, presumably by incorporation into urea, while glutamate so enhanced the rate of ammonia removal that, at 60 minutes, the blood ammonia level was lower than the corresponding level for the controls. They suggested that cirrhotic patients were adapted enzymatically to the formation of glutamine as a substitute for urea synthesis. This is also in keeping with the obser- [illegible] 1955; Velasco and Ducci, 1957).

Our knowledge of ammonia metabolism in liver disease may now be briefly summarized as follows. Ammonia is formed in the gut by bacterial action on urea and proteins and also by enzymatic hydrolysis of amides (Webster and Davidson, 1956). The ammonia is then absorbed into the portal circulation and either by-passes or passes through a damaged liver. It is removed by the peripheral tissues by fixation as carbamyl phosphate and by glutamine synthesis. When these mechanisms become saturated ammonia evolution by the tissues occurs, further elevating the blood level. The toxic action of ammonia might result from either combination with ketoglutarate and depletion of the Krebs cycle or the removal of ATP needed for glutamine synthesis. The consequent shortage of ATP would then result in a failure to maintain ionic gradients across the cell membrane and hence electrical excitability, and a failure of various synthetic functions in the cell already detailed. The mention of ionic gradients and cell permeability in turn raises the question whether ammonia, or to be more precise the ammonium ion, can cross the cell membrane.

pH Changes and Ammonia Toxicity

In 1956 Vanamee, Poppell, Glicksman, Randall and Roberts pointed out that respiratory alkalosis was a common finding in hepatic coma, and they went on to suggest that ammonia stimulated the respiratory centre and that the resultant alkalosis impaired oxygen unloading in the tissues leading to high blood lactic acid levels and diminished cerebral oxygen uptake, which in turn enhanced the toxic action of ammonia. Now these results may be interpreted otherwise, particularly in the light of Warren's studies in mice (Warren, 1958; Warren and Nathan, 1958). He showed that in these animals the toxicity of ammonium salts varied directly with their ability to raise the blood pH and he postulated that this was due to an increased ratio of ammonia gas to the ammonium ion at high blood pHs. Further studies have shown that raising the blood pH increased the speed at which ammonia crossed the blood brain barrier and that the amount of ammonia entering the brain depended more upon the pH than the total amount of ammonia nitrogen administered. In other words as the blood pH approaches the pK for ammonia the ammonia/ammonium ratio shifts towards the unionized gas which can cross the cell membrane readily and accumulate in the cell, where the acid pH rapidly converts it back to ammonium salts. This is a most important observation, but it is a great pity that we do not have any studies made with ammonia at concentrations which bear some relationship to the physiological range.

Now in the light of Warren's (1958) observations the alkalosis of hepatic coma can be re-evaluated. Instead of the raised blood ammonia leading to respiratory alkalosis and reduced oxygen unloading in the tissues, the raised pH allows ammonia to enter the cell and inhibit metabolism resulting in reduced oxygen requirements and hence reduced oxygen uptake.

One link is missing in the chain of evidence needed to incriminate ammonia as the toxic factor in hepatic coma: that is experimental work in animals at ammonia concentrations comparable to those that are found in man. We know that in man, in hepatic coma, the blood ammonia levels vary widely, but seldom exceed 500 µg./100 ml. and are commonly in the range

200–300 µg./100 ml. In no experimental animal is ammonia toxic at such low blood levels. In fact in many animals these figures would be well within the normal range and toxic symptoms are usually encountered only at blood, or tissue, levels around 5,000–9,000 µg. per 100 g., that is about ten times higher than those found in man. Moreover, the toxic symptoms observed are always excitatory and convulsions are a common event. In hepatic coma convulsions are exceedingly rare and the usual picture is one of progressive depression of cerebral function. Until toxic symptoms, similar to those observed in man and at comparable blood levels, can be regularly produced in a suitable experimental animal the evidence for ammonia intoxication in hepatic coma remains inconclusive.

Methionine Toxicity

From ammonia metabolism I now want to turn, briefly, to the question of methionine toxicity. The work of Himsworth and Glynn (1944) and Glynn and Himsworth (1944), who showed that methionine deficiency could cause massive hepatic necrosis in rats, led to the introduction of this aminoacid in the treatment of patients with liver disease. Watson (1949) reported a remarkable patient with cirrhosis who repeatedly lapsed into coma when given methionine, but the significance of this was not at the time appreciated. Later I reported finding methionine sulphoxide in the spinal fluid of a patient in hepatic coma and pointed out that this might block the ammonia binding mechanism in the brain as methionine sulphoxide was known to be a glutamic acid antimetabolite in micro-organisms (Walshe, 1951). In 1955 Sherlock repeated and confirmed Watson's observation on the toxicity of methionine and was later able to show that this did not depend on changes in the blood ammonia (Phear, Sherlock and Summerskill, 1955, Phear, Ruebner, Sherlock and Summerskill, 1956). Almost at the same time Challenger and I (Challenger and Walshe, 1955a, b) reported finding methyl mercaptan and methylsulphide in the urine of a patient in hepatic coma and postulated that these two compounds were probably derived from methionine and discussed their possible toxicity and their relationship to foetor hepaticus.

Dr. J. M. Barnes (1955) studied the toxicity of these compounds for me and found that methyl mercaptan rapidly produced unconsciousness in rats at concentrations as low as 5 mg. per litre in the inspired air.

Moreover, the coma was rapidly reversible on withdrawal of the mercaptan, although the animals continued to smell of the compound for some hours after recovery. The origin of the mercaptan has not been established. Challenger and Walshe (1955b) believed it was formed in the body as a result of deranged methionine metabolism, the enzyme thionase in the liver causing a fission of the carbon-sulphur bond in the methionine molecule and resulting in the formation of methyl mercaptan and homoserine or alpha aminobutyric acid. Phear and others (1956) believed that methionine toxicity was secondary to bacterial action in the gut and they were able to suppress, or retard, the toxic action of orally administered methionine by the use of a broad spectrum antiobiotic. They also reported that the small intestinal juice of patients with cirrhosis was odourless (Martini, Phear, Ruebner and Sherlock, 1957) so presumably the mercaptans formed from methionine were of metabolic origin.

Role of Ammines in Hepatic Coma

Brief mention should also be made of the role of amines in the genesis of hepatic coma. At present there is no direct evidence that toxic concentrations of these compounds are present in the blood of patients in hepatic coma, but intestinal fermentation of protein could well lead to the formation of pharmacologically active amines which would by-pass the liver in the same way as ammonia. One amine which requires mention is serotonin, 5-hydroxy tryptamine; this is a normal body metabolite which may well be concerned with both normal and abnormal brain function.

Bessman, Merlis and Borges (1957) have suggested that a deficiency of this compound might result in a disturbance of consciousness and has given the precursor, 5-hydroxy tryptophan, to patients in coma apparently with improvement in the E.E.G. pattern. The related indolyl compounds have also received attention. As long ago as 1933 Quastel and Wheatley postulated a toxic role for these compounds in hepatic disease

and more recently (1958a) I have shown that a number of indolyl derivatives will inhibit cerebral oxygen uptake and the response to maximal stimulation.

On Susceptibility

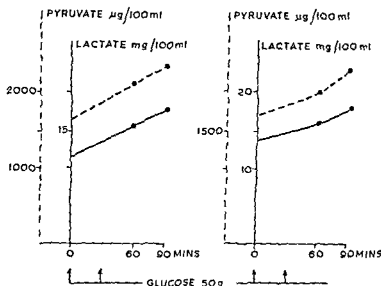
Finally there is one further point requiring consideration. That is the question of susceptibility. We have already seen that both ammonium salts and methionine will induce the syndrome of hepatic coma only in susceptible subjects. These people usually have moderately severe hepatic disease and very often large portal collateral shunts. There are also scattered reports of ammonium toxicity in patients with normal livers, though strictly speaking they can hardly be considered to be metabolically normal, as for instance the patient with dehydration and severe alkalosis described by Wolfe and others in 1958. If we accept this phenomenon of susceptibility, and in view of the available evidence I think we must do so, then it is not unreasonable to say that the brains of patients with liver disease are lacking in some fundamental defence mechanism against intoxication or are deficient in some essential metabolite normally supplied by the liver or a combination of these two factors.

While the work of Geiger and Yamasaki (1956) on the isolated perfused cat's brain has shown that a deficiency of essential factors such as cytidine and uridine can lead to loss of electrical excitability we have no evidence as to whether this occurs in man, but it may well be that there is a shortage of co-enzymes or enzyme activators necessary for normal cerebral energetics.

One such factor which has recently received attention is lipoic acid whose role in pyruvate metabolism has already been discussed. Although there is no evidence that a shortage of this compound occurs in man it has been given, allegedly with benefit, to patients in hepatic coma. However, Bravo, Orrego and Walshe (1959) studied 20 patients with severe hepatic disease who were receiving large doses of thiamin and could find no evidence that treatment with lipoic acid restored disturbed carbohydrate metabolism to normal (Fig. 3).

Similarly inhibition of the normal ammonia binding mechanism might leave the brain vulnerable to concentration of

ammonia which under normal conditions would be harmless. Methionine sulphoxide has already been mentioned as a theoretical possibility here. Other factors are the change in the blood pH allowing ammonia to enter the cell more readily than



Mean of

levels

in the normal and it would be interesting to know the effect of alkalosis on ammonium tolerance of normal individuals. The views of Bessman and Bessman (1955) on ketoglutarate depletion could also be adduced. Chronic low-grade elevation of the blood ammonia level might lead to a slowly progressive depletion of the brain's reserves of ketoglutarate and glutamic acid with an eventual slowing of the Krebs cycle to a point at which it could no longer support consciousness. Thus coma might develop at a blood ammonia concentration apparently

well below the toxic level and, if we invoke the reverse sequence of events, consciousness might not return until the blood level had been low for long enough for the deficiency of essential metabolites to be replaced. As we have seen already it would take some 9 hours for all the glutamate in the brain to be removed at an ammonia extraction rate of 50 µg./100 ml., that is making no allowance for replacement by any mechanism and also assuming that no transamination can occur in the human brain, although we know that this reaction can certainly occur in certain animal species.

In the experimental field there is some evidence to support the theory of susceptibility. Walshe, De Carli and Davidson (1958a, b), in a limited number of experiments on rats with hepatic cirrhosis, showed that the cortex of these animals was able to respond to maximal stimulation less well than normal cortex in the presence of respiratory inhibitors at sub-toxic concentrations. These abnormal cortex slices also produced more ammonia than the normals and were unable to depress their ammonia production on maximal stimulation.

Conclusions

Now let me try to sum up briefly. The following facts are established. The blood ammonia concentration is usually elevated in patients in hepatic coma and there is often an accompanying respiratory alkalosis. The brain may take up ammonia, be in equilibrium, or evolve ammonia into the blood.

A syndrome indistinguishable from hepatic coma can be induced in susceptible subjects by a number of compounds, principally ammonium salts and methionine. The nature of the factor leading to susceptibility has not yet been determined, and until we know more about this we cannot hope to have a fully logical and effective therapy. Indeed, the whole subject is at a rapidly growing point of knowledge and it is hardly surprising that both facts and interpretations are still apparently contradictory. I am sure the patient with impending hepatic coma would join with the poet in saying

> When men of science find out something more
> We shall be happier than we were before.

ammonia which under normal conditions would be harmless. Methionine sulphoxide has already been mentioned as a theoretical possibility here. Other factors are the change in the blood pH allowing ammonia to enter the cell more readily than

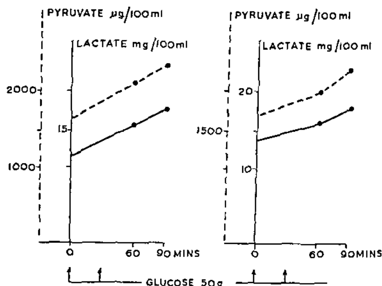


FIG 3 Pyruvate and lactate metabolism in hepatic cirrhosis. Mean of 20 patients with chronic liver disease: 50 g. glucose by mouth at zero time [illegible] results after the patients [illegible] The [illegible] the [illegible] levels.

in the normal and it would be interesting to know the effect of alkalosis on ammonium tolerance of normal individuals. The views of Bessman and Bessman (1955) on ketoglutarate depletion could also be adduced. Chronic low-grade elevation of the blood ammonia level might lead to a slowly progressive depletion of the brain's reserves of ketoglutarate and glutamic acid with an eventual slowing of the Krebs cycle to a point at which it could no longer support consciousness. Thus coma might develop at a blood ammonia concentration apparently

well below the toxic level and, if we invoke the reverse sequence of events, consciousness might not return until the blood level had been low for long enough for the deficiency of essential metabolites to be replaced. As we have seen already it would take some 9 hours for all the glutamate in the brain to be removed at an ammonia extraction rate of 50 μg./100 ml., that is making no allowance for replacement by any mechanism and also assuming that no transamination can occur in the human brain, although we know that this reaction can certainly occur in certain animal species.

In the experimental field there is some evidence to support the theory of susceptibility. Walshe, De Carli and Davidson (1958a, b), in a limited number of experiments on rats with hepatic cirrhosis, showed that the cortex of these animals was able to respond to maximal stimulation less well than normal cortex in the presence of respiratory inhibitors at sub-toxic concentrations. These abnormal cortex slices also produced more ammonia than the normals and were unable to depress their ammonia production on maximal stimulation.

Conclusions

Now let me try to sum up briefly. The following facts are established. The blood ammonia concentration is usually elevated in patients in hepatic coma and there is often an accompanying respiratory alkalosis. The brain may take up ammonia, be in equilibrium, or evolve ammonia into the blood.

A syndrome indistinguishable from hepatic coma can be induced in susceptible subjects by a number of compounds, principally ammonium salts and methionine. The nature of the factor leading to susceptibility has not yet been determined, and until we know more about this we cannot hope to have a fully logical and effective therapy. Indeed, the whole subject is at a rapidly growing point of knowledge and it is hardly surprising that both facts and interpretations are still apparently contradictory. I am sure the patient with impending hepatic coma would join with the poet in saying

> When men of science find out something more
> We shall be happier than we were before.

ACKNOWLEDGEMENT

I am indebted to Dr. J. R. Wilson of Lederle Laboratories for a generous gift of Lipoic Acid.

REFERENCES

BARNES, J. M. (1955). Private communication.

BENITEZ, D., PSCHEIDT, G. R. and STONE, W. E. (1954). *Amer. J. Physiol.* 174, 488.

BESSMAN, S. P. (1956). In *Inorganic Nitrogen Metabolism* Johns Hopkins, Baltimore.

[illegible] *c. exp Biol*, [illegible]

BESSMAN, S. P., SHEAR, S. and FITZGERALD, J. (1957). *New Engl. J. Med.* 256, 941.

BRAVO, M., ORREGO MATTE, H and WALSHE, J. M. (1959) In preparation.

CHALLENGER, F and WALSHE, J. M. (1955a). *Biochem. J.* 59, 372.

CHALLENGER, F. and WALSHE, J M (1955b). *Lancet*, 1, 1237.

COXON, R. V. (1952) *Biochem. Soc. Symp.* 8, 3.

CRANE, R K. and BALL, E G (1951) *J biol. Chem.* 188, 819.

DAWSON, A M., DE GROOTE, J, ROSENTHAL, W. L. and SHERLOCK, S. (1957). *Lancet*, 1, 392.

ELLIOTT, K. A. C. and HENRY, M. (1946). *J biol. Chem.* 163, 361

FAZEKAS, J. F., TICKTIN, H. E, EHRMANTRAUT, W. R and ALMAN, R W. (1956). *Amer J. Med.* 21, 843.

FINDLAY, M, ROSSITER, W. L and MAGEE, R J. (1954) *Biochem J.* 58, 236.

FULD, H. (1933) *Klin Wschr* 12, 1364.

GEIGER, A, DOBKIN, J and MAGNES, J (1953) *Science*, 118, 655

GEIGER, A. and YAMASAKI, S (1956). *J Neurochem* 1, 93.

GLYNN, L. E. and HIMSWORTH, H. P (1944). *J Path. Bact* 56, 297.

GORE, M B R and McILWAIN, H. (1952). *J. Physiol* 117, 471.

HARPUR, R. P and QUASTEL, J H (1949). *Nature, Lond* 164, 693

HIMSWORTH, H P. and GLYNN, L. E (1944). *Clin Sci* 5, 133

HIMWICH, H. E (1951) In *Brain metabolism and cerebral disorders* Williams and Wilkins, Baltimore

KETY, S. S. and SCHMIDT, C F. (1948). *J clin. Invest* 27, 476.

KIRK, E. (1936) *Acta med scand Suppl.* 77.

KREBS, H. A (1935) *Biochem J* 29, 1951.

McILWAIN, H. (1953) *Biochem J* 53, 403

MADDOCKS, S., HAWKINS, J E. and HOLMES, E. (1939). *Amer J. Physiol.* 125, 551.

MANGOLD, R., SOKOLOFF, L., CONNER, E., KLEINERMANN, J, THERMAN, P. G. and KETY, S. S. (1955). *J clin Invest* **34**, 1092
MANN, P. J. G., TENNENBAUM, M and QUASTEL, J. H. (1939). *Biochem J.* **33**, 822.
MARTINI, G. A., PHEAR, E A., RUEBNER, B. and SHERLOCK, S (1957). *Clin Sci.* **16**, 35.
ORTEGA Y GASSET, J. (1951). *Revolt of the Masses* George Allen and Unwin, London.
PHEAR, E A., RUEBNER, B, SHERLOCK, S and SUMMERSKILL, W H. J. (1956). *Clin. Sci.* **15**, 93
PHEAR, E. A, SHERLOCK, S. and SUMMERSKILL, W H. J. (1955) *Lancet*, **i**, 836
J **27**, 1609
5, 1261.
176, 1199
1163.
(1954). *J. clin. Invest.* **33**, 984
SHERLOCK, S (1958). *Amer. J Med.* **24**, 805
SOKOLOFF, L, MANGOLD, R., WECHSLER, R., KENNEDY, C and KETY, S S (1955). *J. clin Invest.* **34**, 1110
J. clin. Invest **36**, 361.
TYOR, M P. and WILSON, W. P. (1958). *J Lab clin. Med* **51**, 592
VANAMEE, P, POPPELL, J. W., GLICKSMAN, A S, RANDALL, H T. and ROBERTS, K E (1956). *Arch intern Med* **97**, 762
VAN CAULAERT, C., DEVILLER, C and HALFF, M (1932a) *C R. Soc. Biol. Paris*, **iii**, 735.
VAN CAULAERT, C, DEVILLER, C and HALFF, M. (1932b). *C R Soc Biol Paris*, **iii**, 739
VAN CAULAERT, C, DEVILLER, C. and HOFSTEIN, J (1932) *C R Soc Biol Paris*, **iii**, 737.
VELASCO, M and DUCCI, H. (1957). *Rev. med Chile*, **85**, 193.
WAELSCH, H. (1951). In *Advances in Protein Chemistry*, vol. vi, Academic Press, N.Y.
WALSHE, J. M. (1951). *Quart J Med.* **20**, 421.
WALSHE, J M. (1953) *Quart J. Med* **22**, 483
WALSHE, J. M. (1955). *Lancet*, **i**, 1235
, C. S. (1958a) *Clin Sci* **17**, 11.
, C. S. (1958b). *Clin Sci* **17**, 27.
7.
J. clin. Invest. **37**, 1724
WATSON, C. J. (1949) *Arch intern. Med* **31**, 405

WEBSTER, L. T. (1956). *Amer. J. Med.* **21**, 130.
WEBSTER, L. T. and DAVIDSON, C. S. (1956). *J. clin Invest.* **35**, 742.
WEBSTER, L. T. and DAVIDSON, C. S. (1957). *J. Lab. clin. Med.* **50**, 1.
WEBSTER, L. T. and GABDUZDA, J. G (1958). *J. Clin. Invest.* **37**, 414.
WEIL MALHERBE, H. (1938). *Biochem. J.* **32**, 2257.
WEIL MALHERBE, H. (1952). *Biochem. Soc. Symp.* **8**, 16.
WEIL MALHERBE, H and GREEN, R. H. (1955). *Biochem. J.* **61**, 210
WHITEHEAD, T. P. and WHITTAKER, S. R. F. (1955). *J. clin. Path.* **8**, 81.
WOLFE, S. F., FAST, B. B, STORMONT, J. M. and DAVIDSON, C. S. (1958) *J. Lab. clin. Med.* **51**, 672.

XXIII

Fine-Structural Studies of the Collagen-Apatite Partnership

W. T. ASTBURY

THE culminating event in the building of bones and teeth is an organized deposition of crystallites of hydroxy-apatite brought about, we believe, through the principal agency of a matrix of collagen fibrils. Except for the crystal structure of apatite, everything in this sentence is vague to some extent, and it is the object of this lecture to try to sharpen the picture a little.

WHAT IS COLLAGEN?

First as regards what we mean by collagen, a name which used to apply to a certain type of connective tissue component only but which has now come to cover a great range of protein fibres, all alike at the fundamental molecular level but often differing widely at constructional levels higher than that: the basic pattern common to them all was revealed in X-ray diffraction studies (1)[1] and is characterized by the now familiar large-angle fibre diagram which in fact still remains the only consistent and generally accepted test for the 'collagens', whatever other functions and properties they may have; for their chemical constitutional scheme has not yet been worked out completely, though it seems that approximately one-third of the amino-acid residues must be glycine residues while another quarter or so consists of proline and hydroxyproline residues, and that the presence of

[1] Since a selection of references only is given in this lecture, serial numbering has been adopted although this differs from the system used elsewhere in the book.

hydroxyproline is unique to the collagens. The molecular structural unit that has been deduced from the large-angle X-ray diffraction diagram (2), supported by electron microscope and physico-chemical studies, has been called 'tropocollagen': it is a stiff, rod-like particle of dimensions some 2,600 Å–2,800 Å by 14 Å and weight about 340,000 (3), made up of three polypeptide chains each twisted into a left-handed helix of three residues per turn, then twisted together into a right-handed 'coiled coil' so that the 'repeat' finally corresponds to the length of ten residues (28·6 Å). This is the image that first comes to mind—this closely integrated, special polypeptide triad—when we think of collagen nowadays, and what we mean by collagen generally is *all* the multitudinous structures that nature has built from it. For the purposes of mineralization, our main business here, tropocollagen is inherently necessary, but it is insufficient in itself, and what is more, since obviously not all collagenous structures have the power of, or are associated with, mineralization, only certain combinations of tropocollagen will do the trick, and even then they may be subject to accessory factors, promoting either activation or inhibition.

Some collagens, for instance the earthworm cuticle, betray no details of their higher organization either by X-rays or in the electron microscope—we see only the large-angle diffraction diagram and unstriated fibrils; but usually a higher periodicity is shown at least in the electron microscope, the common manifestation being about 640 Å, as illustrated most familiarly by the classic example of rat-tail tendon, with its outstanding small-angle diffraction pattern about which we owe so much to Bear and his collaborators, and its particularly evident transverse band pattern that figures so frequently in the extensive electron microscope investigations of the Schmitt and Randall schools and others. There are also found variations on the 640 Å theme: for example, in early stages of fibrogenesis when it can appear as one-third the full period, and in the illustration I have chosen[1] for Plate XXVIII, Figure 1 (collagen fibrils in untreated pulp from a human incisor) where it shows

[1] The lecture itself was illustrated by many more slides than have been selected for this written-up version. Plates referred to will be found between pages 440–1.

itself mostly as a succession of half-periods subdivided again into three; but broadly speaking it is the dominant and 'natural' one in collagen-type structures made *in vivo*, and the implication is straight away that it must somehow be the specific combination of tropocollagen units characterized by the 640 Å periodicity that is responsible in the end—other things being favourable, of course—for the last act of apatite formation. Incidentally, a feature of collagen-type fibrils that cannot be missed in the electron microscope and that must also have its special significance is the way their thickness can differ considerably from one type to another, while within each type remaining extraordinarily uniform. In Plate XXVIII, Figure 1, for example, one notices at once how very much thinner the collagen fibrils in dental pulp are than in rat-tail tendon, say—something like ten times thinner in fact, recalling reticulin, another former 'outsider' that has now been admitted to the collagen family. They are scarcely 300 Å thick and are beautifully uniform, and those seen in the human periodontal membrane are similar.

It was known from the early work of Nageotte, Fauré-Fremiet, Wyckoff and Corey, and others that native collagen can be dissolved in dilute acetic acid and then be reprecipitated essentially unchanged by addition of salts at appropriate pH, but of recent years this finding has been shown, largely at the hands of Schmitt and Gross (4) and their collaborators, to be only a beginning, a first indication of the possibilities: no fewer than five principal forms (and intermediate varieties) of reconstituted collagen are now recognized and their interchangeability *in vitro* established beyond question. The quickest summary of our present knowledge on this score is conveyed by the Schmitt-Gross diagram reproduced in Figure 2, and it tells at a glance the sort and scope of problem we are concerned with, at the very least, in seeking to identify what fundamental combination or combinations of tropocollagen are capable of taking part in mineralization.

And before proceeding further, perhaps a few words are called for on the necessity to distinguish, as a mental reservation at any rate in case there are independent steps in the manufacture and laying-down of apatite in bones and teeth, between

initiation (activation, nucleation) and organization (orientation). The situation is analogous in the first place to 'staining' collagen by various means, such as by phosphotungstic acid to display in the electron microscope the intra-band fine structure,

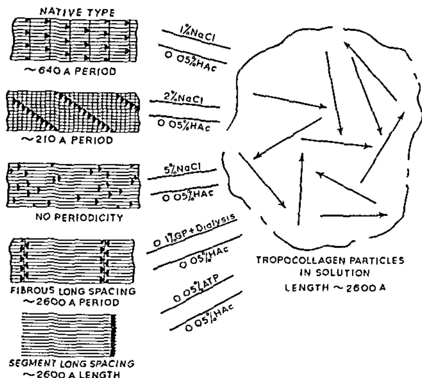


Fig 2

Fig. 2. Modes of aggregation of tropocollagen (after Schmitt and Gross)

which is an expression of more or less specific interactions with amino-acid residues following one another in a presumably definite order, but which need not therewith involve *subsequent* crystal growth and preferred orientation. Actually, phosphotungstic staining of collagen, if overdone, does suggest also this second stage, as shown very strikingly in Plate XXIX, Figure 3, an electron micrograph obtained by Millard of rat-tail tendon fibrils after excessive treatment with PTA. I doubt whether I can

improve on this illustration, if compared with the more usual pictures of PTA-stained collagen, to draw attention in passing to the point—rather obvious, but none the less important to keep in mind—that we have to explain not merely the chemical-biological formation of apatite but also its final remarkable systems of oriented crystallization.

In a similar context it is worth mentioning also the recent findings of Schwarz (5) on staining collagen with silver. He has concluded that the affinity for silver is associated with the main dark band (Bear's less ordered region), and that it is developed through a broad 'spectrum' of connective tissue fibres, deposition taking place in random fashion on the *surface* of embryonic sclera for instance, on the *surface*, too, of reticulin fibrils but in topographical relation to the dark band, and eventually *inside* the dark band with collagen fibrils. He states that all embryonic and newly forming connective tissues take this path in their differentiation, though they do not all reach the stage of intrafibrillar staining, stopping short each at its own particular function; going beyond, indeed, leading to impairment of function, as when the cornea develops opacity. Likewise, we have Fitton Jackson's (6) observations on osteogenic tissue in embryonic fowl. She describes how at an early stage of calcification dense particles, less than 100 Å in size, become localized in a ring in the major region of indentation, that is in the main interband (Bear's more ordered region), of the collagen fibrils; more precisely, they are seen almost exclusively between the so-called *d* and *ab* bands. Electron diffraction diagrams indicate at the same time that they are hydroxyapatite crystallites, *not preferentially oriented*, though, with respect to the fibril axis, in contrast to what has been found with adult bone by Engstrom (7) and his co-workers, using X-ray diffraction. Their crystallites were elongated, about 220 Å by 73 Å, and also tended to lie parallel to the collagen fibre axis; as a consequence of which they put forward their well-known proposal that the numerical connecting link is one-third the 640 Å period.

Tropocollagen Units and Mineralization

But to return now to the question of what manifestation, what special mode of aggregation, of the tropocollagen units is a primary requisite for hydroxyapatite formation, the most impressive recent experiments are those of Glimcher (8) and collaborators, who have made metastable solutions of calcium and phosphate ions and have tested the action on them of the dispersed tropocollagen macromolecules themselves and of all their principal *in vitro* combinations—native-type fibrils (about 640 Å repeat), fibrils of about 220 Å repeat, unstructured fibrils, and structures of 'fibrous long-spacing' and 'segment long-spacing' (both about 2,600 Å repeat). Of these forms only the native-type fibrils were found to have the capacity of inducing the formation of hydroxyapatite crystallites, identified positively by X-ray and electron diffraction and by chemical analytical data. Moreover, it did not matter from what source, whether active *in vivo* or not, the 640 Å fibrils had been reconstituted; for example, those obtained from rat-tail tendon worked, though collagen fibres freshly dissected from this tendon had no power of calcification under similar physico-chemical conditions even when ground up to destroy possible diffusion barriers. Carefully decalcified bone collagen worked, however, without being submitted to the reconstitution process. To round off the series, too, there was the complementary demonstration that no mineralization was initiated if the 640 Å pattern was disorganized by heat, alkali, etc.; and, as a side-test with a high-period, striated, non-collagenous protein, negative results were obtained with paramyosin fibrils from adductor muscles of clams.

Thus the hint from the biological collagens, that it is specifically the 640 Å combination of collagen units that functions in bone and tooth formation, appears to be substantiated; and further than that, just as we could say that tropocollagen alone was necessary but insufficient, so we may conclude now that the 640 Å manifestation *in vivo* is in general *more then sufficient.* Contrary to the commoner view held hitherto, an inhibitor must usually be present—possibly chondroitin sulphate—which is extracted in the reconstitution experiments just described. It is not there in the system operating in normal biological

mineralization, and should it be absent also from tissues where it ought to be present, then the inference is that this could lead to pathological mineralization.

In these studies, as in those of Fitton Jackson, again no preferential orientation of the hydroxyapatite crystallites was detected, and Glimcher brings forward reasons for believing that orientation in bone, for instance, is a directed growth imposed on initiation nuclei formed *inside* the collagen fibrils—trapped, as it were, between closely packed tropocollagen protofibrils—an interpretation from which it would follow that there is no inevitable association between initiation and organization. Plate XXX, Figure 4, reproduces an electron micrograph kindly sent me by Dr. Glimcher for this lecture. It is of reconstituted, highly purified collagen fibrils (at a magnification of 74,000 ×) during the very early stages of *in vitro* mineralization, and the preparation was unstained and unshadowed. The inset is a selected-area electron diffraction diagram of the crystallites seen as small black dots, showing the typical reflections of apatite. These crystallites are unoriented—there is no evident tendency for their *c*-axes to lie parallel to the fibril axes—but they appear to be nucleated at specific sites along the fibril, the majority being spaced about 640 Å apart, that is one per axial period.

Crystallite Growth in Bone, Dentine and Enamel

There is a curious gap in the argument, as has been pointed out by Pautard, after the small rounded hydroxyapatite particles reported by Fitton Jackson and Glimcher as first forming at apparently a single site within the 640 Å period, because the electron micrographs of Engstrom and Fernández-Morán (9) and others of mature bone reveal continuous chains of crystallites elongated in the direction of the collagen fibres (and roughly of the bone axis) and mostly of a rather uniform length approximately one-third 640 Å. Analogous structures to those in mature bone are typical of dentine too, and in this regard I should like now to refer briefly to some investigations carried out by Pautard (10) and Millard (11) in our laboratory at Leeds. They have used a new technique for overcoming in a way the difficulty of sectioning hard mineralized tissues, by

embedding *powders* of dentine and enamel in polymerized mixtures of butyl and methyl methacrylates and cutting them with simply a glass knife. Many of the particles are smashed, but a surprising number, of dentine at any rate, are sectioned neatly enough, as illustrated in Plate XXXI, Figure 5; and at higher magnifications, as in Plate XXXII, Figure 6, the overall result is a fine-structural panorama. The predominant appearance in dentine is of fusiform particles arranged end to end, often in parallel arrays running in the same general direction, but also frequently swirled round into labyrinths. Two or more particles in a row, sometimes straight and sometimes bent, are common, and when they are separated their tapered shape is clear. In transverse section hexagonally packed groups occasionally occur. Recalling again the observations of Fitton Jackson and Glimcher, these strings of crystallites seem, however, to start out as rows of small rounded bodies, which eventually elongate and join up as units similar in dimensions (of the order of 200 Å × 60 Å, according to Pautard) to those found not only in mature bone but also in the ossicles of the protozoon *Spirostomum ambiguum*, which we shall consider below.

The marked fibrous layout of the chains of crystallites in the dentine sections strongly suggests that they are the mineralized expression of a corresponding arrangement of collagen filaments. Nothing of the sort can be seen directly, but that is not unexpected, since it is well known that it is unusual, either by X-ray diffraction or in the electron microscope, to be able to detect collagen in undecalcified dentine preparations, though it can be made immediately obvious in larger pieces, by both methods, once the apatite has been dissolved away. Experiments of the kind are in progress to try to uncover recognizable collagen filaments even in these thin sections. They are almost certainly there but, to judge by the minute size of the apatite units formed in association with them, probably completely encasing them, they may be no more than two or three protofilaments thick.

Sectioning enamel powders by the new method turned out to be not such a clean-cut business, so to speak, as it had proved with dentine, but it had its advantages in that it broke down the

structure into individual crystallites and stacks of crystallites that were full of interesting features—see Plate XXXIV, Figure 9, for example. Electron diffraction diagrams, by the way, of selected areas of the electron microscope field bring out at once

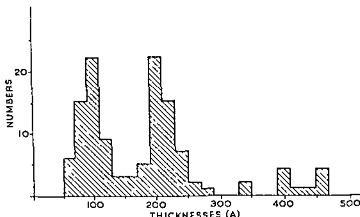


FIG 8

FIG 8. Frequency-distribution of thicknesses of apatite crystallites in an electron micrograph of enamel powder (Pautard).

that the fundamental crystalline units in enamel are much larger than those in dentine, as will be seen from Plate XXXIII, Figures 7a and 7b. They are very thin elongated platelets (some 400–1,000 Å wide and several 1,000 Å long) of which the basic thickness, as inferred from frequency-distribution measurements carried out by Pautard on a shadowed preparation (Fig. 8), is only about 100 Å, and they are stacked like cards in a pack.

The most fascinating feature of these enamel platelets is the rather regular arrays of circular holes or thinner places, having thicker or denser peripheries and spaced roughly 100 Å apart, with which they are so frequently decorated, while in the background appear at the same time collections of faint rings, like quoits. There is often seen, too, what looks like a counterpart manifestation: corresponding arrays of particles on the surfaces of the crystallites, with many similar particles in the background as if they had once fitted the holes or depressions and had been dislodged, or alternatively were nuclei capable of side-to-side

co-operation in building crystallites but which never got a proper start; the 'quoits', in fact, suggest 'ghosts' of such incipient crystallites. Altogether, the impression is that each thin platelet was initially 'threaded on', and grew out sideways from, a collection of filaments, stacks of platelets being built up by the superposition of these elementary sheets formed at successive active levels. If there is anything in this idea, it would mean that the relation between the collagen-apatite partnership in dentine and in enamel is a sort of complementary perpendicular one, periodic sites along the filaments initiating the process in both cases, and crystallites then growing out longitudinally in dentine but laterally in enamel, the latter being morphologically more perfect besides. Another idea, catalysed by the occurrence in some of the dentine sections of sequences of particles spaced not at about 200 Å but at about 100 Å apart, is that the 'magic number' which would link up bone and dentine *and* enamel is the latter: witness the thickness of the enamel platelets and the spacing of the depressions and particles on their surfaces, the similar dimensions of the initial particles described by both Fitton Jackson and Glimcher, and the sub-period along the collagen fibrils in dental pulp and the periodontal membrane.

MINERALIZATION IN A PROTOZOON

From the advanced collagen-apatite partnership in bone and tooth formation I should like before concluding to glance back for a moment at something that might well exemplify their first tentative acquaintance in the simpler organisms—investigations we are carrying out at Leeds following on the discovery that hydroxyapatite crystallites could be deposited in close association with algal flagella (*Polytoma*) in certain culture conditions, notably involving bactotryptone and tap water (12). Pautard (13) then obtained X-ray evidence for the presence of bone salts in the protozoon *Spirostomum ambiguum*, and in due course found in the electron microscope, and identified as hydroxyapatite by X-ray diffraction, 'ossicles' like those shown sectioned in Plate XXXV, Figure 10. This alone is intriguing enough, but Figure 10 should be compared with Plate XXXVI, Figure 11, an electron micrograph (Pautard and Millard) of 'ossicles' in the

femur of a two-months' cow foetus. The resemblance is striking; and furthermore, it seems that these structures too are both built up from elongated units of much the same dimensions as those in adult bone and dentine. Pautard is therefore of the opinion that such units, if not exactly universal, are probably widespread in biological apatite formation, with the implication, of course, that they arise from a common type of organic template of the nature or potentialities of collagen. No collagen, for instance by X-ray diffraction tests, has been detected so far in *Spirostomum*, but that is no discouragement either, in an obscure evolutionary field only just being opened up, which I personally find most stimulating. There are plenty of possible lines of continuation that could lead to what, as molecular biologists, we are all searching for, the precise circumstances in which collagen or a precursor was first co-opted into the affair of mineralization. One of the prime objects of mineralization is to provide mechanical support for muscular activity, and in this connection it is extremely suggestive that *Spirostomum* becomes more and more 'ossified' with age and the necessity for actual burrowing for food. Pautard puts forward the theory that dynamic existence requires stored phosphate for the prosecution of active phosphorylation processes, and that this is particularly the case in a muscular, boring animal, over and above the mechanical aspect.

Conclusion

Finally, in my 'manifestation' as crystallographer, perhaps I may be allowed to dream a little about some future fine-structural lecture in this series. The lecture I visualize may still be a long way off in time, but all the same, even from the present sketchy story, I hope you will agree that the outlines are already beginning to show. What we glimpse now is the framework of the fundamental collagen macromolecule—minus a lot of important details, it is true, but they will come; a first indication of how nature combines this unit into a great range of fibrils of many functions, and of which particular combination it is that is involved in biological mineralization; evidence for hydroxyapatite nucleation, at least, at specific sites in the periodic 640 Å pattern along this chosen fibril; and further evidence that

co-operation in building crystallites but which never got a proper start; the 'quoits', in fact, suggest 'ghosts' of such incipient crystallites. Altogether, the impression is that each thin platelet was initially 'threaded on', and grew out sideways from, a collection of filaments, stacks of platelets being built up by the superposition of these elementary sheets formed at successive active levels. If there is anything in this idea, it would mean that the relation between the collagen-apatite partnership in dentine and in enamel is a sort of complementary perpendicular one, periodic sites along the filaments initiating the process in both cases, and crystallites then growing out longitudinally in dentine but laterally in enamel, the latter being morphologically more perfect besides. Another idea, catalysed by the occurrence in some of the dentine sections of sequences of particles spaced not at about 200 Å but at about 100 Å apart, is that the 'magic number' which would link up bone and dentine *and* enamel is the latter: witness the thickness of the enamel platelets and the spacing of the depressions and particles on their surfaces, the similar dimensions of the initial particles described by both Fitton Jackson and Glimcher, and the sub-period along the collagen fibrils in dental pulp and the periodontal membrane.

Mineralization in a Protozoon

From the advanced collagen-apatite partnership in bone and tooth formation I should like before concluding to glance back for a moment at something that might well exemplify their first tentative acquaintance in the simpler organisms—investigations we are carrying out at Leeds following on the discovery that hydroxyapatite crystallites could be deposited in close association with algal flagella (*Polytoma*) in certain culture conditions, notably involving bactotryptone and tap water (12). Pautard (13) then obtained X-ray evidence for the presence of bone salts in the protozoon *Spirostomum ambiguum*, and in due course found in the electron microscope, and identified as hydroxyapatite by X-ray diffraction, 'ossicles' like those shown sectioned in Plate XXXV, Figure 10. This alone is intriguing enough, but Figure 10 should be compared with Plate XXXVI, Figure 11, an electron micrograph (Pautard and Millard) of 'ossicles' in the

PLATE XXVIII



FIG 1 A field of collagen fibrils in untreated pulp from a human incisor

oriented crystal growth then proceeds along the directions of the fibrils or in simple relation thereto, and that the building stone is similar in dentine and bone, approximately matching in length one-third the 640 Å period, but takes a form in enamel that looks intrinsically different but could still be only another geometrical expression of the same ruling principles. We could represent these findings in a general way on this occasion by means of a model of patterned strings running in various preferred directions with rows of crystallites threaded on them also in preferred orientations with respect to the strings and fitting on to the string pattern, and it might all seem pretty vague, as I said; but it is not so very vague, in my view. I believe such a model would already contain the gist of the matter; and as Wagner is reported to have remarked when he had completed writing the words of *The Ring*, we have only now to write the music.

Writing the music consists of tracing the paths and defining the groupings of the fibrils, perfecting the amino-acid analysis [illegible] this leads to its known modes [illegible] t which side-chains in their [illegible] complex interact with which components of the known lattice structure of hydroxyapatite—all this at a most modest estimate. It is a tall order, as the saying goes, but it can be done. And the dream-model I see in my mind's eye for this dental research lecture in the, I trust, not too distant future could be a molecular-architectural masterpiece of the kind foreshadowed so inspiringly in the X-ray investigations of Thewlis (14), for example, on the crystallographic organization of hydroxyapatite in human deciduous enamel. Each enamel prism is composed of crystallites which fall into two groups, so arranged that their hexagonal axes tend to make angles of approximately 5° and 40° with the direction of the prism; and the crystallite axes point away from the tooth to that side of the prism remote from the nearest cusp. That is the class of 'blueprint' we have to work to, and it is largely a ques- [illegible] structure [illegible] on for [illegible]



FIG 4 Electron micrograph of reconstituted highly purified collagen fibrils during the very early stages of *in vitro* mineralization Preparation unstained and unshadowed magnification $\times 74{,}000$ The inset is a selected-area electron diffraction diagram (apatite) of the crystallites seen as black dots (Glimcher)



FIG 3 Electron micrograph of collagen fibrils in rat-tail tendon heavily stained with phosphotungstic acid (Millard)

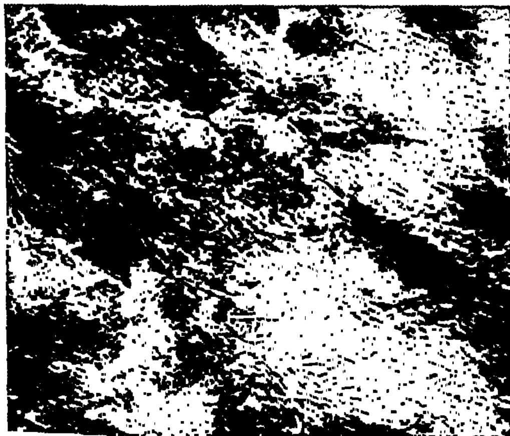


Fig 6 As in Fig 5 but at high magnification



FIG 5 Low-magnification electron micrograph of a thin section of dentine particles in a powder ground from a human immature first molar



Fig. 9 Electron micrograph of thin plate-like crystallites in a sectioned powder of enamel from a human immature first molar

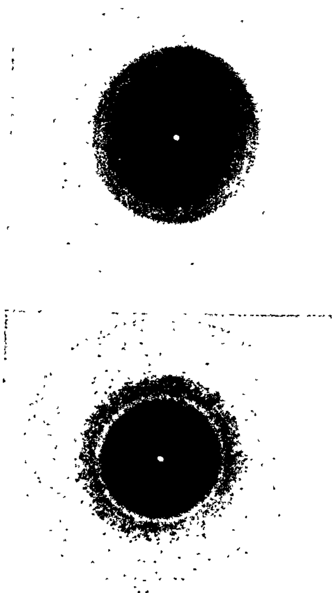


FIG 7 (*a, upper*) Electron diffraction diagram of an electron microscope field of sectioned dentine powder from a human immature first molar (*cf* Figs 5 & 6), (*b, lower*) ditto, enamel (*cf* Fig 9).



FIG. 9. Electron micrograph of thin plate-like crystallites in a sectioned powder of enamel from a human immature first molar



FIG. 10. Electron micrograph of sectioned hydroxyapatite 'needles' ...

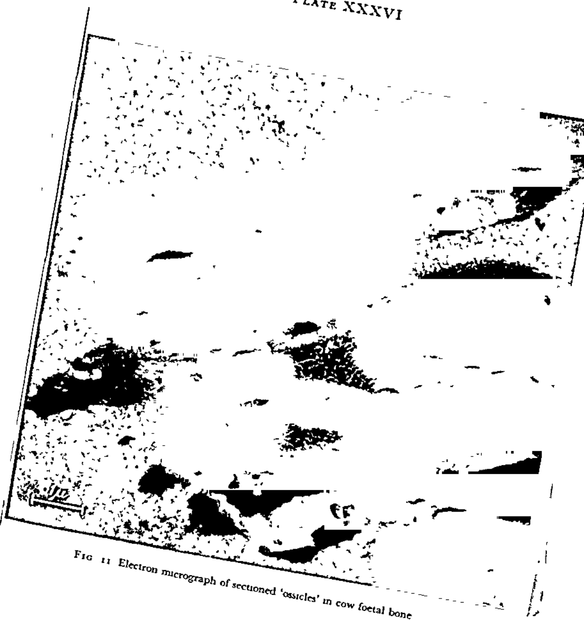


FIG 11 Electron micrograph of sectioned 'ossicles' in cow foetal bone

ACKNOWLEDGEMENTS

I am indebted to Mr. Millard for all the electron micrographs reproduced here, with the exception of Fig. 4, which I owe to Dr. Glimcher. I wish to thank also Dr. Pautard for the benefit of many discussions.

REFERENCES

(which will also serve as a guide to many others)

1. W. T. ASTBURY, *Trans Faraday Soc.* **34**, 378 (1938), R. S. BEAR, *Adv. Protein Chem.* **7**, 69 (1952).
2. [illegible]
176, 1062 (1955).
3. H. BOEDTKER and P. DOTY, *J Amer. Chem. Soc.* **77**, 248 (1955).
4. F. O. SCHMITT, *Proc. Amer Phil Soc.* **100**, 476 (1956); J. GROSS, *J. biophys. biochem. Cytology*, **2**, 261 (1956).
5. W. SCHWARZ, *Connective Tissue* a symposium organized by the Council for International Organizations of Medical Sciences, p. 144 (1957). (Blackwell Scientific Publications, Oxford)
6. S. FITTON JACKSON, *Proc. Roy Soc. B*, **146**, 270 (1957).
7. D CARLSTROM, A. ENGSTROM and J. B. FINEAN, *Symp Soc. exptl Biol.*, **9**, 85 (1955).
8. M. J GLIMCHER, A. J. HODGE and F. O. SCHMITT, *Proc. Nat Acad. Sci.* **43**, 860 (1957), M J. GLIMCHER, *Rev. mod. Phys* **31** (2), 359 (1959).
9. H FERNÁNDEZ-MORÁN and A. ENGSTROM, *Biochim Biophys. Acta*, **23**, 261 (1957).
10. F G E PAUTARD [illegible] Conference Manchester [illegible]
11. [illegible]
12. [illegible]
13. [illegible] *Acta*, **28**. 514 (1958), XV th Internat. Congr Zool., Sect VI, paper 54 (in press), *Biochim. Biophys Acta*, **35**, 33 (1959). A.A.A.S Symposium on Calcification in Biological Systems, Dec. 1959 (in press).
14. J. THEWLIS, *Proc. Roy. Soc* B, **127**, 211 (1939).

XXIV

The Biochemical Background to the Action of Fluoride in Dental Caries

G. NEIL JENKINS

THE fact that dental caries incidence is reduced in districts with water supplies which contain about 1 p.p.m. of fluoride is one of the best authenticated facts in dental research. Although there can be no serious doubt about the existence of an anti-caries effect of fluoride, its mode of action is still controversial. Of the two main theories, (1) that fluoride lowers the solubility of enamel or (2) that it inhibits bacterial enzymes, opinion is crystallizing in favour of the first, but as I hope to show, the whole question is hedged in by doubts and difficulties. Before discussing in detail the evidence for these two theories I will briefly mention the form in which fluoride is believed to be present in the hard tissues.

Fluoride may combine with the mineral matter of bone and tooth in several ways. When exposed to concentrations of more than 75 p.p.m. of fluoride, hydroxyapatite forms calcium fluoride which adsorbs on to the crystal but can be fairly readily washed off. Some fluoride may exchange with hydroxyl groups to form fluorapatite in which the fluoride is irreversibly bound. At fluoride concentrations below 75 p.p.m. this last reaction is probably the only one to occur (Leach, 1959). It seems clear that from the *levels of fluoride in physiological fluids* (*about* 0 1 p.p.m.) fluoride can only enter the calcified tissues as fluorapatite. After the 'topical application' of fluoride, i.e. painting the tooth with a 2 per cent solution of sodium fluoride, calcium fluoride will mainly be formed, which explains why the

fluoride is rapidly removed from most parts of the tooth within a week or so of its application (Zwemer, 1957).

In considering the possibility that fluoride might leave the enamel surface and enter the dental plaque in concentrations high enough to inhibit the bacteria, Speirs and I argued that if this happened, then the fluoride on the surface of enamel might be expected to be in lower concentration than throughout the enamel as a whole. We therefore compared the fluoride of inner and outer enamel. The outer enamel was studied by exposing the crowns of intact teeth to several changes of acetic acid and estimating the fluoride in the acid washings. Phosphate was also estimated in order to measure the amount of enamel dissolved. The rest of the enamel was separated from the dentine by standard methods and its fluoride estimated. The results (Table 1) showed very clearly that in teeth from both high and low fluoride areas the surface fluoride concentration was several times *higher* than that of the enamel as a whole and that all the absolute figures were related to the fluoride intake from the water supply (Jenkins and Speirs, 1953). It was shown that even unerupted teeth possessed this surface fluoride layer. When we studied deciduous teeth by this method we could not find a fluoride-rich layer on the surface but other methods have since shown it to be present.

TABLE 1. Fluoride concentration of surface and inner enamel (Jenkins and Speirs, 1953)

Source of teeth	Fluoride of water supply (p p m)	F p p m Surface	F p p m Interior
West Hartlepool	2 0	1,310	270
South Shields	1 4	960	110
North Shields	<0 25	580	80

Brudevold, Gardner and Smith (1956), and Isaac, Brudevold, Smith and Gardner (1958b) confirmed and extended these results by estimating the fluoride in successive layers of enamel *ground* off the tooth. They found that the fluoride concentration increased with age both on the surface and within the enamel, although they confirmed our finding that the general pattern of

distribution exists even in unerupted teeth. In teeth from subjects with 5 p.p.m. of fluoride in their water supplies the surface *concentration* did not increase with age although there was a movement inward of fluoride, as the fluoride-rich layer was found to become thicker. It is important to note that the general trend is for more fluoride to be picked up by enamel, although, contrary to what we believed when this work began, I think that this finding does not necessarily rule out the temporary loss of some fluoride into the plaque.

There is good evidence that the main effect of fluoride is exerted through structural effects on the teeth. Most of the effect depends on the ingestion of fluoride during enamel formation although there is undoubted evidence of some effect after eruption. Also, if people bred in a high fluoride area move out of it, they still show an increased resistance to caries.

The Solubility Theory of Fluoride Action

Since fluorapatite is less soluble than hydroxyapatite an obvious possible way in which fluoride might reduce caries is by reducing the solubility of the tooth surface.

The theory would seem to be inherently probable since, as already mentioned, the fluoride concentration on the enamel surface (like the caries incidence) varies with the intake—at least over the range which is relevant to the anti-caries action. One would expect an inverse relation between fluoride content and solubility although there are few data on this particular point and at least one worker (Rathje, 1952), in experiments that appear not to have been repeated, found that after a certain concentration was reached further increases in fluoride have no effect on solubility.

The truth of this theory has often been assumed on the strength of demonstrations of the big differences between the solubilities of ground enamel before and after treatment with concentrated sodium fluoride solution. Surprisingly, there has been very little experimental work in which the solubility of the normal outer surfaces of teeth from high and low fluoride areas have been compared. Since this information seemed to me to provide the crucial test of the theory, I began some years ago

to collect figures for the solubility of teeth from natives of North and South Shields. When the work began these two very similar towns, separated only by the few hundred yards of the River Tyne, provided an almost ideal testing ground for the fluoride effect; North Shields had a fluoride-free water supply, South Shields had about 1·4 p.p.m. and Weaver (1944) had *shown the dramatic effect on caries incidence*. As the work progressed, however, South Shields took on additional sources of water, some being low in fluoride, so that it now averages only about 0·5 p.p.m. over the whole town and varies from part to part. It became necessary to abandon South Shields as a source of experimental material and the comparison was continued between North Shields and West Hartlepool where the water contained about 2 p.p.m. and where caries was lower even than in South Shields (Weaver, 1950).

The solubility was studied by Brudevold's method in which teeth, covered in wax except for a window of 4 mm. diameter, are shaken for ten minutes in acetic acid buffer at pH 4·0 and the phosphate dissolved is estimated. Brudevold found, and other users of the method have confirmed, that the solubility of enamel varies enormously from tooth to tooth and even in different parts of the same tooth. It is therefore necessary to study large numbers to establish small differences and it is not easy to collect adequate numbers of comparable teeth. It is particularly difficult to obtain supplies of teeth from high fluoride areas because the lower incidence of caries makes fewer extractions necessary. The results showed a consistent trend towards a lower solubility in the deciduous teeth from the high fluoride areas but this was only statistically significant for deciduous teeth from North Shields and West Hartlepool (Table 2). An even smaller difference was detected in permanent teeth from adults (Table 4). Speirs and I had the idea of comparing the solubility of the powdered enamel ground off the outer layers of groups of teeth varying in fluoride content. We thought this technique would overcome the individual variation and facilitate the detection of small differences, but the results of a few preliminary experiments were unpromising so this approach was dropped Isaac, Brudevold, Smith and

Gardner (1958a), however, reported with this technique differences of up to 5 per cent in the solubility (as measured by weight loss in acid) of enamel powder ground off teeth from high and

TABLE 2. Solubility of deciduous teeth from areas with 0, 1 and 2 p.p.m. fluoride (Jenkins *et al.* (1952) and unpublished)

Fluoride of water supply	North Shields 0	South Shields 1	North Shields 0	West Hartlepool 2
P dissolved (µg./ml.)	10·4	9 2	12·0	9·5
	diff. 11 5% not sig.		diff 21% sig.	
Number of teeth	77	79	108	109

Layer	0 1 p.p.m. F F content (p p.m.)	0 1 p.p.m. F % wt. loss in acid	1 p.p m. F F content (p.p m.)	1 p.p m. F % wt. loss in acid	% diff. in solubility
1	1,080	35·7	1,552	32·8	2·9
3	305	44·4	328	42·7	1·7
5	198	48 8	219	43 3	5·5

low fluoride areas (Table 3). The solubility of each succeeding layer rose as the fluoride content fell. Nevertheless, they found that some samples (e.g. layer 5 in the table) which had similar [illegible]

finding that although the fluoride concentration of the enamel rose with age, there was little change in solubility, for example one batch from an 'over 50' age group contained 1,080 p.p.m. of fluoride and dissolved more rapidly than another from the 'under 20' group with only 499 p.p.m. In view of results like this, the positive differences on solubility must be interpreted with caution. It was presumably effects of this sort which confused

our own attempts to compare the solubilities of powdered surface enamel from North Shields and West Hartlepool.

Another inherent difficulty in measuring the solubility of tooth substance has also emerged from these studies. When enamel powder is treated with acid so that some of it dissolves, it has been found that the fluoride concentration of the undissolved residue is higher than the original. Apparently, as the apatite dissolves, the fluoride ions released into the solution become reattached to the new surface which is exposed as the dissolved layer is removed. During a solubility test, therefore, the enamel becomes progressively less like the original material. The effect of this on the solubility figure is difficult to assess but obviously it must tend to confuse the picture. It is likely that reattachment of fluoride from the acid washings on to the enamel explained our failure to detect a fluoride-rich layer on the surface of deciduous teeth. It is known that all fluoride values are lower in deciduous compared with corresponding permanent teeth and the tendency for reattachment of fluoride might be expected to be greater in those surfaces least saturated with fluoride. It is, in fact, remarkable that we detected fluoride in the washings even of permanent teeth; Isaac and his coworkers (1958a) found little or none in the acids after exposure to their enamel samples.

Solubility of Enamel in Early Cavities

Myers, Hamilton and Becks (1952) treated the enamel surface with ^{18}F and showed that there was a slight uptake over the surface as a whole, but a much greater uptake in damaged enamel areas including early carious cavities. These results have been confirmed by Hardwick, Fremlin and Mathieson (1958). This raised the possibility that in the normal mouth, fluoride from saliva, food and drink might become concentrated in very early cavities and other enamel defects. If this happened to a greater extent in high fluoride areas then any influence which fluoride might have on solubility would presumably be greater in the fluoride area. Dowse and I therefore studied the fluoride content of enamel ground out of early cavities and the solubility of the surfaces of early carious lesions from teeth in high and low

Gardner (1958a), however, reported with this technique differences of up to 5 per cent in the solubility (as measured by weight loss in acid) of enamel powder ground off teeth from high and

TABLE 2. Solubility of deciduous teeth from areas with 0, 1 and 2 p.p.m. fluoride (Jenkins *et al.* (1952) and unpublished)

Fluoride of water supply	North Shields 0	South Shields 1	North Shields 0	West Hartlepool 2
P dissolved (μg./ml.)	10·4	9·2	12 0	9·5
	diff. 11·5% not sig.		diff. 21% sig.	
Number of teeth	77	79	108	109

[illegible]

Layer	0·1 p.p.m. F — F content (p.p.m.)	% wt loss in acid	1 p.p.m. F — F content (p.p.m.)	% wt loss in acid	% diff. in solubility
1	1,080	35·7	1,552	32 8	2 9
3	305	44·4	328	42·7	1 7
5	198	48 8	219	43·3	5 5

low fluoride areas (Table 3). The solubility of each succeeding layer rose as the fluoride content fell. Nevertheless, they found that some samples (e.g. layer 5 in the table) which had similar fluoride concentrations did not always have the same solubility, i.e. fluoride is apparently not the only constituent of teeth affecting its solubility. This was illustrated further by their finding that although the fluoride concentration of the enamel rose with age, there was little change in solubility, for example one batch from an 'over 50' age group contained 1,080 p.p.m. of fluoride and dissolved more rapidly than another from the 'under 20' group with only 499 p.p.m. In view of results like this, the positive differences on solubility must be interpreted with caution. It was presumably effects of this sort which confused

as enolase (Warburg and Christian, 1942) and the cytochrome system (Borei, 1945) and partly from the finding that in high fluoride areas the salivary lactobacillus counts tend to be lower. The low lactobacillus counts by no means prove an antibacterial effect, however, since there is increasing evidence that these organisms are the *result* of open carious lesions rather than their cause. At any rate, Becks (1950) and Kesel and his colleagues (1958) have found that the lactobacillus counts in patients with rampant caries fell markedly after their cavities were filled.

In order to assess this theory information is required on three points. First, we must know the minimum concentration of fluoride necessary to inhibit salivary bacteria; secondly, we must know the minimum degree of inhibition which will have a significant effect on caries; and thirdly, we must know the concentration of fluoride present in the dental plaque under various conditions. Conclusive information is lacking on all these points but we are especially ignorant about the third.

The Minimum Inhibitory Concentration of Fluoride

Bibby and van Kesteren (1940) and Bibby, Volker and van Kesteren (1942) studied the effect of a range of concentrations of sodium fluoride on the acid production and growth of pure cultures of various strains of salivary organisms. Their results showed that 2 p.p.m. of sodium fluoride had some effect on acid production but that much higher levels (up to 500 p.p.m.) were required to reduce growth. Wright and Jenkins (1954) carried out similar experiments on acid production in saliva and found that 0 5 p.p.m. F was the lowest concentration which produced a statistically significant inhibition.

Borei (1945) reviewed the factors which have been found to influence the inhibitory effect of fluoride and emphasized the following: pH at which the organisms are exposed to fluoride, state of nutrition of the organisms (whether fed or starving), and the magnesium and phosphate concentration of the medium. Warburg and Christian (1942) have explained the action of magnesium and phosphate by finding that a magnesium fluorophosphate complex is the real inhibitor of enolase and its formation is proportional to the concentration of each of these ions

fluoride areas. The results (Dowse and Jenkins, 1957) showed that early cavities contain, throughout their depth, more than double the fluoride concentration of comparable amounts of sound enamel ground out from the same tooth and—although, as I have emphasized, all solubility studies must be accepted with caution—the solubility of the surface of lesions in high fluoride areas was significantly less than in low fluoride areas (Table 4). The experiments were also of interest, first in showing

TABLE 4. Solubility and fluoride content of carious and non-carious parts of permanent teeth from areas with 0 and 2 p.p m. fluoride in water (Dowse and Jenkins, unpublished data)

	Non-carious part		Carious part	
	0 p p.m.	2 p p.m.	0 p.p m.	2 p p m.
P dissolving	12·75	12·1	8·2	6·7
	Diff. 5% not sig.		Diff. 18% sig.	
F (p p.m.)	250	440	640	950
Number of teeth	108	112	108	112

that even in a non-fluoride area the inorganic matter of an early lesion is only two-thirds as soluble as the neighbouring intact enamel, and secondly in confirming that the differences in solubility of intact surfaces of enamel from adults in high and low fluoride areas shows only non-significant trends as measured by this technique.

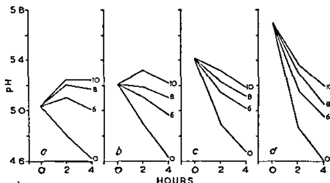
These results might lead to the cautious suggestion that the effect which fluoride has been found to have after eruption is to combine with the enamel in early lesions, presumably owing to its greater permeability, and to slow the progress of the caries.

Summarizing the evidence as a whole on the solubility theory we may say that it tends to support it, but for various technical reasons it is by no means consistent or conclusive.

THE ANTIBACTERIAL THEORY

I turn now to the main rival of the solubility theory—the antibacterial or antienzyme theory. This arose partly from the well-known fact that fluoride does poison certain enzymes such

the concentrations completely prevent acid production although they still have a marked effect. The cause of the rise in pH in the presence of fluoride calls for some comment. I have not yet investigated this fully, but there are indications that it



(By courtesy of *Arch oral Biol.* and Pergamon Press.)

FIG. 1 Effect of adding 0, 6, 8, 10 p p.m. of fluoride to saliva-glucose mixtures adjusted to initial pH values of (*a*) 5 04, (*b*) 5 2, (*c*) 5 4, (*d*) 5 7.

occurs simply because when acid production is completely inhibited the alkali-producing (and acid-destroying?) mechanisms of salivary bacteria are unopposed.[1]

These observations raised the possibility that even when added at neutrality to saliva-glucose incubation mixtures, the effect of fluoride does not begin until the pH reaches values below 6·0. To test this point, saliva containing glucose was incubated with and without fluoride from an initial pH of 7 0 and samples were withdrawn periodically so that the point at which inhibition began could be detected. It was found (Fig. 2) that although some inhibition occurs at quite high pH values the main divergence in the presence of the fluoride is below 6 0. I think these results, along with the absence of phosphate in his medium, largely explains why Lilienthal found, in most of his experiments, that fluoride concentrations below 19 p.p.m. did not inhibit—his medium was buffered at 6 8 which meant that

[1] Experiments carried out since this paper was read have shown that this conclusion is incorrect Fluoride has since been found to enhance slightly those reactions of salivary bacteria which raise the pH

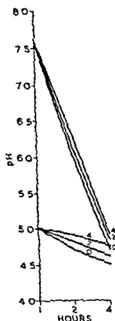
present. Lilienthal (1956) was the first to investigate these points from the point of view of the anti-caries action of fluoride. He studied the effect of a range of concentrations of fluoride under various circumstances on the anaerobic acid production (measured as carbon dioxide) of salivary sediment suspended in bicarbonate buffer at pH 6·8. The medium was in many respects unphysiological and the method precluded the possibility of investigating the effect of pH on the fluoride inhibition. His results showed that in most experiments the minimum inhibitory concentration was between 19 and 32 p.p.m. He found that calcium prevented fluoride inhibition, presumably by precipitation of calcium fluoride which would readily occur at pH value of 6·8, and he suggested that the calcium of saliva would have a similar effect. When phosphate was added to the medium, however, an inhibition was found with 0·5 and 1·0 p.p.m. fluoride. In spite of this last result, which agreed with our findings on saliva (where phosphate is, of course, present), Lilienthal did not accept our results as valid and suggested that the differences we found were within experimental error.

We think it more probable that our results, referring to a natural saliva environment containing magnesium and phosphate ions and in which calcium is soluble, provide more reliable data on the minimum effective inhibition. My own unpublished observations confirm Lilienthal's finding that the state of nutrition of salivary bacteria does not affect their fluoride sensitivity.

THE EFFECT OF pH ON FLUORIDE INHIBITION

Recently I [illegible]
tion of saliv[illegible]
to various [illegible]
centrations of 0, 6, 8 and 10 p.p.m. Glucose was added and the mixtures were incubated for 4 hours. The results (Fig. 1) clearly show that as the pH approaches 5·0 there is a sharp increase in sensitivity to fluoride. At 5·0, all three levels of fluoride completely inhibit acid production for 4 hours and the pH actually rises. When the initial pH was 5·2 only 10 p.p.m. inhibited acid production completely whereas at 5·7 none of

The finding that 8 p.p.m. of fluoride can be a powerful inhibitor shows, incidentally, that the precipitation of calcium fluoride is unlikely to be a limiting factor to fluoride action, since the solubility at neutrality is 16 p.p.m. (≡8 p.p.m. F),

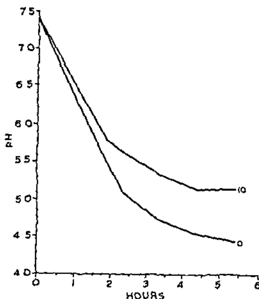


(By courtesy of *Arch oral Biol* and Pergamon Press)

FIG. 3 The effect of 0, 2 and 4 p.p.m. of fluoride added to saliva-glucose mixtures at pH values of 7 6 and 5·0.

i.e. even if precipitation is occurring, the concentration of ions in solution would still be adequate for inhibition. If the ionic concentrations are such that calcium phosphate is precipitated, however, the fluoride might become bound to it by ionic exchange leaving only very low concentrations in solution This probably occurred in Lilienthal's experiments in which he found, at pH 6 8, that even 38 p.p.m. of fluoride had no inhibitory effect in the presence of 2×10^{-3} *M* Ca^{++} and 5×10^{-3} *M* PO_4'''.

It is concluded that the concentration of 0·5 p.p.m. found by



(By courtesy of *Arch oral Biol* and Pergamon Press.)

Fig. 2. Serial pH readings on saliva-glucose mixtures incubated with and without 10 p.p.m fluoride from an initial pH of 7 5.

the organisms never reached their state of maximum sensitivity. A direct comparison of the effects of 2 and 4 p.p.m. of fluoride on saliva acid production from different initial pH values (Fig. 3) confirmed the general conclusion that in saliva, very low concentrations inhibit and that the pH at which the organisms are first exposed to the fluoride influences their sensitivity. Although most of these experiments on the effect of pH were carried out on saliva which was adjusted to a value of 5·0, the results have been confirmed on saliva which reached this pH value by normal bacterial metabolism of sugar, and which were, therefore, in a more physiological condition (Jenkins, 1959a).

The action of additional magnesium and phosphate ions, which have been found in yeast to enhance the fluoride effect, have not been thoroughly studied in saliva, but unpublished preliminary experiments have failed to show any consistent effect, presumably because saliva already contains these ions.

THE CONCENTRATION AND SOURCE OF FLUORIDE IN THE PLAQUE

The final point to discuss is the concentration of fluoride present in the plaque. This is at present a matter for speculation but one which may be solved by the methods of Hardwick and others (1957) with radioactive fluoride.

There are three possible sources of fluoride in the plaque. The first is the saliva which is presumably the usual liquid medium of this material. Saliva has been found to have a fluoride concentration averaging about 0·1 p.p.m. From all the published data, this level of fluoride is too low to influence bacterial metabolism significantly (but not too low to be a source from which enamel uptake could occur) and, in any case, it has not been found to vary with the level in the drinking water. So little is known about the composition of plaque that the possibility that fluoride is concentrated by it cannot be dismissed.

The second possible source of plaque fluoride is drinks. It is probable that in places the plaque becomes soaked with fluids during drinking, and in a fluoride area or with strong tea the fluoride concentration of the drink would be about 1–2 p.p.m. The plaque concentration would presumably be lower owing to dilution by the fluid already present (probably saliva and therefore low in fluoride). Unless the plaque can bind fluoride, the drinks would have to wash over the plaque at the same time as, or very shortly after, food is taken so that the fluoride is present as acid production occurs. If one had the appropriate muscular actions during drinking to spread the fluid and if one drank at the right time in relation to food, then drinks might have some influence in raising plaque fluoride concentrations up to between 1–2 p.p.m. which, at least *in vitro*, does reduce acid production significantly. On the whole, however, a direct effect of fluoride in drinks on plaque bacteria does not seem a likely mechanism.

The third possible source of fluoride is the enamel surface. The work of Bibby and Van Kesteren (1940) was planned to test the possibility that fluorosed enamel exerts antibacterial effects and although they came to a positive conclusion, their results are capable of quite a different interpretation (see Jenkins,

Wright and Jenkins to be the minimum for significant inhibition is confirmed and that under some conditions, as little as 6 to 8 p.p.m. can stop acid production altogether.

The Minimum Inhibition which will affect Caries

There is no direct evidence on the question of how much inhibition of acid production is required to have a significant effect on caries, but there are grounds for believing that an inhibition which is small in absolute terms might be of importance. Saliva at neutrality is saturated with calcium phosphate (Ericsson, 1949) and consequently enamel does not dissolve. When the pH falls, the concentrations of calcium and phosphate ions become less and less able to maintain the saturation, so that at some value, usually between 5·5 and 6·5, the enamel begins to dissolve. Consequently small differences of pH near the critical value can have an important influence on the amount of enamel dissolving. To take an extreme case as an example, if the critical pH of a particular saliva is 5·5 and the plaque pH falls to 5·4, then some enamel will dissolve. If, however, fluoride or some other inhibitor were to prevent the pH from falling below 5·5 (a very small absolute difference—only 0·1 of a pH unit) it would make the difference between some decalcification and none at all. Similar considerations apply to changes in solubility—a small difference might have a profound effect. This concept of a critical pH may require revision if Schatz's so- [illegible] This theory [illegible] constituents [illegible] citrate ions, may dissolve the enamel above the critical pH (Schatz *et al.*, 1958). This idea is plausible enough but no one seems to have shown yet that salivary organisms can produce chelators which could attack the enamel on a scale comparable with acid decalcification. There have, however, been few investigations of this point. Some preliminary experiments of my own (Jenkins, 1959b) were negative, but this does not of course disprove the theory.

apatite falls, so that the fluoride would have its maximum tendency to recombine with apatite immediately after release—and if this happened to all of it none would remain free to inhibit the organisms. Nevertheless, some might reasonably be expected to enter the cells, because of the concentration gradient, and thus become unavailable for immediate reattachment by the enamel. Although the tendency for the uptake becomes less with rising pH it is still quite large at 6·0 (Neuman and Neuman) but has stopped by pH 9·0; they give no data for pH values between pH 6·0 and 9·0. There is thus no data which would contradict the possibility that a good deal of the fluoride could become reattached although the process suggested above would undoubtedly tend to lead to some fall since it is unlikely that *all* the fluoride would become reattached; some would diffuse out of the plaque, and some would be removed during tooth-brushing and other forms of plaque removal. The observed fact is, of course, that enamel fluoride rises with age. Since fluoride can be taken up by enamel even from very low concentrations, this gradual rise could conceivably occur by uptake from saliva between meals.

So far, I have discussed the antibacterial theory in terms of reduced acid production in the presence of fluoride but another possible mechanism exists. The exact site at which fluoride inhibits the metabolism of salivary bacteria is not known, but by analogy with its effects elsewhere, it would be expected to interfere with enolase, the enzyme which converts phosphoglyceric acid into phosphopyruvic acid in the Embden-Meyerhof-Parnas scheme. If this is true it would be expected that phosphoglyceric acid would accumulate in the plaque at the expense of lactic acid. I have reinvestigated the suggestion of Osborn (1941) that certain organic phosphates reduce the acid solubility of tooth substance, and found it to be correct (Jenkins, 1957). If, therefore, fluoride led to a mixture of phosphoglyceric and lactic acids being formed in the plaque, this mixture would, I believe, be less effective in dissolving enamel than would the same pH produced by lactic acid alone. I have not been able to show that phosphoglyceric does in fact accumulate in saliva inhibited by low levels of fluoride and Fosdick (1939) had

Armstrong and Speirs, 1952). Nevertheless, the inner plaque is in contact with enamel which, even in the comparatively coarse layers which can be ground off, may contain over 1,000 p.p.m. of fluoride on a 'high fluoride' tooth. In the outermost few molecules of apatite the concentration is probably much higher. This fluoride is bound irreversibly to the enamel (Leach, 1959) so the only way by which enamel fluoride would enter the plaque is by dissolving off the apatite crystals. When the plaque pH falls below the critical figure we can imagine the innermost parts being flooded with ions dissolving off the surface—not only calcium and phosphate but magnesium and fluoride would enter an acidified plaque. It would not seem impossible that at least in part of the plaque the concentration might, from a 'high fluoride' tooth, reach the 6–8 p.p.m. which, my experiments suggest, would stop further acid production completely. Even lower concentrations would have a marked effect in reducing further fall in pH and in cutting down further decalcification. It is still possible that the phosphate and magnesium also entering the plaque might favour fluoride inhibition; the experiments on this point were not encouraging but more work is needed. The enamel surface would seem to be the only source of concentrations of fluoride high enough to have a really marked effect on acid production. It is realized that this process will not be confined to 'high fluoride' teeth (since all teeth possess the fluoride-rich layer), but it would be expected to be most marked in those teeth with the highest concentration of fluoride on their surface.

Since the fluoride level of enamel rises with age it must be supposed that any fluoride dissolved off in the way suggested above must either become reattached or be replaced. As the plaque pH rises some time after a meal (partly by the outward diffusion of the acid, partly by inward diffusion of saliva buffered about neutrality and partly by exhaustion of carbohydrate substrate), conditions would develop which would favour the precipitation of some of the calcium phosphate and the reattachment of the fluoride. Unfortunately for the theory, Neuman and Neuman (1958) and Leach (1959) have shown that as the pH rises, the tendency for fluoride to bind with

Kesel, R. G., Shklair, I. L., Green, G. H. and Englander, H. R. (1958). *J dent. Res.* **37**, 50.
Myers, H. M., Hamilton, J. G. and Becks, H (1952) *J dent Res* **31**, 743.
Neuman, W. F. and Neuman, M. W. (1958). *The Chemical Dynamics of Bone Mineral.* University of Chicago Press, Chicago.
Osborn, T. W. B. (1941). *J. dent. Res.* **20**, 59.
Rathje, W. (1952). *J. dent Res* **31**, 761.
Schatz, A, Martin, J J and Schatz, V. (1958). *Rev. belge de science dentaire*, **13**, 538.
Warburg, O. and Christian, W. (1942). *Biochem. Z.* **310**, 384.
Weaver, R. (1944). *Brit. dent. J.* **76**, 29.
Weaver, R. (1950). *Brit. dent. J.* **88**, 231.
Wright, D. E. and Jenkins, G. N. (1954) *Brit. dent. J.* **96**, 30.
Zwemer, J. D. (1957). *J. dent Res.* **36**, 182.

difficulty in detecting it even with very high concentrations of fluoride. Also, it is possible that any phosphoglycerate formed by the plaque bacteria would remain in the cells where it could not affect enamel solubility. The importance of this speculation remains in doubt, therefore.

Although it is clear that the whole idea of an antibacterial theory of fluoride action is speculative I have given my reasons for believing that the virtual rejection of this theory by Lilienthal and Martin (1956) and Leach (1959) seems hardly justified. Only more facts, especially information on the fluoride concentration in plaque, can settle this complicated problem.

REFERENCES

BECKS, H. (1950). *J. Calif. Dent. Ass.* 26, 53.

BIBBY, B. G. and VAN KESTEREN, M. (1940). *J dent. Res* 19, 391.

BIBBY, B. G., VOLKER, J. F. and VAN KESTEREN, M. (1942). *J. dent. Res.* 21, 61.

[illegible] *dent. Res.* 35, 420.

DOWSE, C. M. and JENKINS, G. N. (1957). *J. dent. Res.* 36, 816.

ERICSSON, Y. (1949). *Acta odont. scand* 8, Supplement 3.

FOSDICK, L S (1939). *J. Amer dent. Ass.* 26, 415.

HARDWICK, J. L., FREMLIN, J. H. and MATHIESON, J. (1958). *Brit dent. J.* 104, 47.

ISAAC, S, BRUDEVOLD, F, SMITH, F. A. and GARDNER, D E (1958a) *J. dent. Res* 37, 254.

ISAAC, S., BRUDEVOLD, F, SMITH, F. A and GARDNER, D E. (1958b). *J. dent Res.* 37, 318

[illegible] ORCA Conference, Lund [illegible]

[illegible] (1952). *Proc. Roy Soc Med* 45, 517.

[illegible] *J Physiol.* 121, 21P.

[illegible] *J. dent. Res.* 35, 189

L. C. THOMSON
*The Physiological Basis of Visual Sensation

J. M. MACKINTOSH
*The Contribution of Science to the Practice of Health in the First Quarter of the Twentieth Century

ALICE M. STEWART
Methods of Research in Social Medicine

G. P. CROWDEN
Environmental Factors in Work

DONALD HUNTER
Methods of Research in Industrial Medicine

H. D. KAY
*Recent Light on Mammary Function

E. C. DODDS
*Research on Ageing

A. HADDOW
Carcinogenesis

W. V. MAYNEORD
Physical Techniques in the Medical Application of Ionizing Radiations

J. F. LOUTIT
*Biological Effects of Radiation

O. G. EDHOLM
*The Effects of Haemorrhage on the Cardiovascular System in Man

M. L. ROSENHEIM
*Lability of Blood Pressure

J. M. MCMICHAEL
Cardiac Output in Man (*film*)

1952–53

J. Z. YOUNG
The Influence of Language on Medicine

SIR FREDERICK BARTLETT
The Nature and Place of Thinking in Medicine

SIR JAMES SPENCE
*The Methodology of Clinical Science

BRADFORD HILL
The Statistical Approach

C. G. DOUGLAS
Control of Respiration

SIR B. H. C. MATTHEWS
Life at High Altitudes

W. K. STEWART
*The Physiological Effects of Gravity

F. J. W. ROUGHTON
The Kinetics of Rapid Chemical and Biological Reactions

D. KEILIN
Metal Catalysis and Intracellular Respiration

J. T. RANDALL
Biophysical Studies of Connective Tissue

A. H. T. ROBB-SMITH
*The Functional Significance of Connective Tissue

E. J. KING
*Silicosis

MATTHEW STEWART
Pulmonary Asbestosis

HONOR B. FELL
Organ Culture in Biological and Medical Research

P. C. C. GARNHAM
*The Life History of the Malaria Parasite

J. C. WHITE
*Human Haemoglobins

SIR ALAN DRURY
Uses and Applications of Human Blood Plasma Fractions

A. S. PARKES
*Preservation of Living Cells at Low Temperatures

P. L. MOLLISON
*The Life-span of Red Blood-cells

A. NEUBERGER
*Biochemical Genetics

C. E. DENT
*Chromatography in the Study of Amino-Acid Metabolism

ADRIAN ALBERT
*Selective Toxicity (*2 lectures*)

COMPLETE LIST OF LECTURES

* *Lectures included in the published volumes*

1951–52

E. ADRIAN
*The Scientific Approach to Medical Research

H. DINGLE
The Philosophy of Science

W. T. ASTBURY
Studies by X-Ray Analysis, Electron Microscopy and Supporting Techniques of the Structure of Biological Macromolecules and the Tissues formed from them

G R CAMERON
*Tissue Responses to Injury

D. VAN SLYKE
*Studies of Normal and Pathological Physiology of the Kidney

R. A. PETERS
Biochemical Function of Vitamin B_1

S. J COWELL
Nutritional Science and the Feeding of Populations

J YUDKIN
*Nutritional Assessment of the Individual

D D WOODS
Folic Acid and Vitamin B_{12} in the Metabolism of Micro-organisms

B C. J. G. KNIGHT
Aspects of Bacteriostasis

A. A. MILES
*Some Aspects of Antibacterial Immunity

A. W. DOWNIE
*Antibodies and Immunity to Virus Infection

F. C. BAWDEN
Current Knowledge on Nature of Viruses

S. P. BEDSON
*Viruses as the Causes of Diseases

F G. YOUNG
*Adrenal Hormones and ACTH

SIR ROBERT ROBINSON
Chemical Aspects of Antibiotics

J M TANNER
*Growth of the Human at the Time of Adolescence

W D. NEWCOMB
Bone Growth and Repair

L J WITTS
The Rate and Materials of Blood Formation

L J. WITTS
Alimentary Factors in Blood Formation

R. G. MACFARLANE
*Blood Coagulation in Theory and Practice

SIR CYRIL BURT
The Psychology of Personality

SIR GEOFFREY JEFFERSON
*On the Organization of Cortical Mechanisms

G L BROWN
Involuntary Nervous System (*2 lectures*)

LORD STAMP
*The Action of Bacterial Enzymes on Immunizing Antigens

H. B. MAITLAND
*Antiviral Immunity

C. C. CHESTERMAN
*Antimalarial Drugs

F. M. D'ARCY HART
Chemotherapy of Tuberculosis

R. R. RACE
Blood Groups

J. V. DACIE
Congenital Haemolytic Anaemia

J. H. KELLGREN
*The Supporting System and its Disorders

G. M. BULL
*Regulation of Body Water

1954–55

A. V. HILL
*Why Biophysics?

B. S. PLATT
*Protein Malnutrition

J. D. BOYD
The Importance of Detail in the Clinical Anatomy of the Autonomic Nervous System

D. WHITTERIDGE
*The Effects of Visceral Distension

P. M. DANIEL
Some Features of the Peripheral Circulation and Vascular Bed

W. T. J. MORGAN
*The Chemical Basis of Blood Group Specificity in Man

D. M. WILSON
*The Electrolyte and Metabolic Response to Trauma

J. D. JUDAH
*Enzymes in Injury

J. P. BULL
*Shock from Burns

A. G. EVERSON PEARSE
*Histochemistry and its Application to the Basic Sciences

SIR VICTOR E. NEGUS
*Comparative Anatomy of the Larynx

R. I. S. BAYLISS
Factors Influencing Adrenocortical Activity in Health and Disease

L. MARY PICKFORD
*Release and Action of Posterior Pituitary Hormones

ROSALIND PITT-RIVERS
Thyroid Hormones

F. BERGEL
*Some Chemical Aspects of Abnorma Growth

A. HADDOW
Theory and Application of the Nitrogen Mustards

SIR RUDOLPH PETERS
Medical Significance of Biochemical Lesions

DOUGLAS McCLEAN
*Substances that Increase Tissue Permeability and their Relation to Infection and Fertilization

D. C. MELROSE
*Cooling of the Whole Organism

JANET VAUGHAN
*Radiation Effects on Bone

E. F. GALE
Actions of Antibiotics on Bacteria

P. B. MEDAWAR
Tissue Transplantation Immunity

J. R. SQUIRE
*Correlation between Laboratory and Clinical Findings in Hypersensitivity

C. H. GRAY
*The Chemistry of the Porphyrias

T. S. WORK
Protein Biosynthesis

J. N. DAVIDSON
*Nucleoproteins in Cell Structure

J. H. GADDUM
*The Effects of Alcohol

A. C. FRAZER
*Fat Metabolism

SIR ALEXANDER FLEMING
*Recent Progress in Antibiotics

WILSON SMITH
*Virus Adaptability in Relation to Human Disease

F. R. PEACOCK
*Carcinogenesis

G. F. MARRIAN
*The Metabolism of the Adrenocortical Hormones

S. J. FOLLEY
The Pituitary Gland and Reproduction

A. ST. G. HUGGETT
*The Physiology of Parturition

W. S. FELDBERG
*The Physiology of the Autonomic Nervous System

W. D. M. PATON
*The Principles of Ganglionic Block

R H S THOMPSON
*Cholinesterases and Anti-Cholinesterases

H J SEDDON
Certain Aspects of Nerve Repair

J. D. BOYD
Development of the Heart in Relation to Congenital Heart Disease

A HEMINGWAY
Dynamics of the Heart Beat

G W PICKERING
The Natural History of Essential Hypertension

E. P SHARPEY-SCHAFER
Causes of Hypotension in Man

1953–54

SIR HENRY DALE
Scientific Method in Medical Research

R. W. RUSSELL
*Experimental Psychopathology

G POPJÁK
*Biological Synthesis

J H BURN
*Acetylcholine and the Maintenance of the Cardiac Rhythm

F. G. YOUNG
*The Growth Hormone of the Anterior Pituitary Gland

E. C AMOROSO
The Biology of the Foetal Membranes and the Placenta

P. M. F. BISHOP
*The Physiological Actions of the Sex Hormones

E. J. KING
*Acid and Alkaline Phosphatase in Disease

R L. M. SYNGE
Principles of Chromatography

W. T ASTBURY
Some Recent Ideas about the Proteins and other Biological Macromolecules

G. W. HARRIS
*Stress and Thyroid Activity

H. L SHEEHAN
The Vascular Lesions of Pituitary Necrosis

H. E. SIGERIST
*Science and History

W D M. PATON
*Anticholinesterases

R. M B MACKENNA
*The Scientific Approach to Dermatology

E A CARMICHAEL
*Hemispherectomy and the Localization of Function

G R CAMERON
*Tissue Repair

D. D REID
The Design of Clinical Experiments

J B. S HALDANE
*The Genetics of Some Biochemical Abnormalities

E BOYLAND
*The Chemotherapy of Cancer

L. P GARROD
*Causes of Failure in Antibiotic Therapy

A. A MILES
*Reactions to Bacterial Invasion

F. H. CRICK
The Structure and Function of Small Viruses

V. P WHITTAKER
*The Metabolism of Choline Esters

MARY BARBER
*Resistance of Staphylococci to Antibiotics

E. C. AMOROSO
Endocrinology of Pregnancy

W. J HAMILTON
Multiple Pregnancy

W J. DEMPSTER
*Homotransplantation of Organs

C G. ROB
*Arterial Substitutes

SIR RUSSELL BROCK
The Relation between the Embryology and Morphology of the Heart, and Cardiac Surgery

P WOOD
The Physiological Basis of Clinical Signs in Heart Disease

J. C GILSON
*Tests of Lung Functions: an Assessment of their Usefulness

D. G. MELROSE
*Principles of Heart-Lung Machines

C LONG
*Phospholipids

SIR SOLLY ZUCKERMAN
Biological Systems

G HADFIELD
*The Role of the Endocrine System in Breast Cancer

A J LEWIS
*Social Psychiatry

E. G. T. LIDDELL
*Cajal and Sherrington

H McILWAIN
*Neurochemistry

E. J ZAIMIS
*Factors Influencing the Action of Neuromuscular Blocking Substances

C. A. KEELE
*Causes of Pain

N. B. MYANT
*Biliary Excretion of Thyroid Hormone

H M. SINCLAIR
*Vitamins in Nutrition

H. J. SEDDON
*Recent Work on Paralytic Poliomyelitis

W. J. WHELAN
*The Synthesis and Degradation of Polysaccharides

A. L. GREENBAUM
*The Control of Fat Metabolism

HONOR B. FELL
*The Physiology of Skeletal Tissue in Culture

J B. DUGUID
Arterial Occlusion

H. GRUNEBERG
The Localization of Inherited Disease in the Body

1957–58

J. HUXLEY
Some Biological Aspects of Cancer

W. S. PEART
*Some Biochemical Aspects of Hypertension

G L. MONTGOMERY
*Some Problems in the Pathology of Coronary Artery Disease

J P. SHILLINGFORD
*The Study of the Circulation by Dye Dilution Curves

ROSEMARY BIGGS
*Haemophilia and Christmas Disease

J. V DACIE
*Acquired Haemolytic Anaemia

H LEHMANN
Variations in Haemoglobin Synthesis

D L MOLLIN
*The Megaloblastic Anaemias

R. A. KEKWICK
*The Plasma Proteins

C. H. ANDREWES
*The New Look in Virus Research

1955–56

SIR WILFRID LE GROS CLARK
*Hypothesis and Speculation in Scientific Research

M. M. SWANN
Physiology of Mitosis

H. G. DAVIES
*The *Use of the Interference Microscope in Biological Research*

O. G. EDHOLM
*The Effects of Cold on Man

A. C. DORNHORST
**Physiology of the Lower Oesophagus and Cardia*

R. E. TUNBRIDGE
*Observations on the Structure of Connective Tissue Fibres

N. H. MARTIN
*Some Aspects of Protein Diseases

R. D. HARKNESS
*Metabolism of Collagen

J. C. MCCLURE BROWNE
*Isotopes in the Study of Problems of Pregnancy

G. S. DAWES
*Physiological Effects of Anoxia in the Foetal and Newborn Lamb

J. WALKER
*The Oxygen Environment of the Foetus

D. A. SLOME
*Physiology of Nasal Circulation

AUDREY U. SMITH
*Experimental Hypothermia in Animals

J. N. HUNT
*The Investigation of Gastric Digestive Function in Man

R. E. DAVIES
Biochemical Aspects of Gastric Secretion

H. A. KREBS
*Steering of Metabolic Processes

W. F. J. CUTHBERTSON
*The Nutrition of Micro-Organisms

R. A. MORTON
*Vitamin A

K. M. RUDALL
*Protein Ribbons and Sheets

M. D. MILNE
*Renal Control of Acid-Base Balance

D. R. WILKIE
*Living Muscle

S. V. PERRY
**Proteins in Muscular Contraction*

J. P. STOKES
*The Treatment of Hepatic Coma

LORD ROTHSCHILD
Fertilization

T. R. R. MANN
Mammalian Semen. Composition and Function

E. S. HORNING
*Some Anomalies in Endocrine Carcinogenesis

K. C. RICHARDSON
Problem of Mammary Growth and Structure

M. W. GOLDBLATT
**Industrial Toxicology*

J. M. BARNES
*The Elucidation of Toxicity

J. A. V. BATES
*Observations on the Cortical Motor Areas in Man

J. F. LOUTIT
*Recovery from Lethal Effects of Ionizing Radiation

1956–57

SIR JOHN COCKCROFT
*Biological Significance of Atomic Energy

K. M. SMITH
*The Electron Microscope in the Study of Viruses

COMPLETE LIST OF LECTURES

A. D. M. GREENFIELD
*The Regulation of the Blood Vessels in the Limbs

C. V. HARRISON
*Pathology of Pulmonary Hypertension

F. B. BYROM
*The Significance of Hypertensive Encephalopathy

O. M. WRONG
*Sodium Excretion and the Control of Extracellular Fluid Volume

A. F. ROGERS
Human Physiology in Antarctica

J. M. WALSHE
*Biochemical Studies in Hepatic Coma

P. FATT
Synaptic Excitation and Inhibition in the Central Nervous System

L. REID
*Chronic Bronchitis and Hypersecretion of Mucus

R. H. MOLE
*Radiation as a Toxic Agent

T. A. J. PRANKERD
*Viability and Survival of Red Cells

A. A. MILES
*Mediators of the Vascular Phenomena of Inflammation

D. F. CAPPELL
*Iron Metabolism

G. H. LATHE
Metabolism of Bilirubin

C. L. COPE
*The Measurement of Adrenal Activity in Man

Scientific Basis of Dentistry

W. T. ASTBURY
*Fine-structural Studies of the Collagen-Apatite Partnership

H. A. SISSONS
The Structure of Calcified Tissues

G. N. JENKINS
*The Biochemical Background to the Action of Fluoride in Dental Caries

H. O. SCHILD
Mechanism of Anaphylaxis

J. A. B. GRAY
*Peripheral Mechanisms Underlying Sensations

J. H. HUMPHREY
*Antibody Metabolism

G. GORDON
*Central Sensory Representation

H. J. TAYLOR
Physiological Problems Associated with Diving and Underwater Swimming

K. W. CROSS
*Anoxia of the New-born

R. J. W. REES
*Experimental Approach to the Problems of Resistance to Tuberculosis

L. S. PENROSE
*Biochemical Genetics and Medicine

L. E. GLYNN
*The Structure and Stability of Collagen in Relation to Diseases of Connective Tissue

G. PAYLING WRIGHT
*Cell Regeneration and Pathological Processes

J. A. FRASER ROBERTS
*Blood-Group Genetics

L. H. GRAY
*The Influence of Oxygen on the Response of Cells and Tissues to Ionizing Radiation

A. F. HUXLEY
Excitation and Contraction in Voluntary Muscle

EDITH BÜLBRING
*Physiology and Pharmacology of Intestinal Smooth Muscle

M. G. P. STOKER
*Growth of Viruses

C. S. HALLPIKE
*Remarks upon the Scientific Basis of Otological Practice

J. M. YOFFEY
*Some Problems of Lymphoid Tissue

A. NEUBERGER
Enzymes in Health and Disease

P. FOURMAN
*Potassium Depletion

P. L. KROHN
*Processes of Ageing in the Reproductive System

C. H. STUART-HARRIS
Adenoviruses and Respiratory Disease in Man

G. W. A. DICK
*Poliomyelitis

1958–59

E. SALISBURY
*The Influence of Botany on Health and Disease

C. H. STUART-HARRIS
*Adenoviruses and Respiratory Disease in Man

D. G. EVANS
*Vaccination against Poliomyelitis

G. S. WILSON
Infectious Enteritis

R. E. O. WILLIAMS
*Epidemic Staphylococcal Infection in Hospitals

W. HAYES
*Bacterial Genetics and Gene Structure

J. H. HUMPHREY
*What are Gamma Globulins?

E. H. MERCER
*Electron Microscopy and the Structure of the Living Cell

M. ABERCROMBIE
*Cell Surface and the Control of Growth

C. N. ARMSTRONG
*Intersexuality

E. F. SCOWEN
*Hyperoxaluria

M. MAIZELS
*The Place of Chemical Pathology in Medicine

J. H. DIBLE
The Pathology of *Limb Ischaemia*